

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 15:30:58 ; Search time 202 Seconds
(without alignments)
5558.757 Million cell updates/sec

Title: US-09-674-593-10

Perfect score: 246

Sequence: 1 gtcattaaaccttgcaggag.....ttcatatcaaccacacacta 246

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 2998549 seqs, 2282253817 residues

Total number of hits satisfying chosen parameters: 5997098

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA.*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	38.4	15.6	861	15	US-10-198-846-1685
C 2	36.2	14.7	3588	13	US-10-282-122A-16553
C 3	35.6	14.5	507	9	US-09-879-536-826
C 4	35.6	14.5	24387	10	US-09-764-891-9844
C 5	35.4	14.4	2803	9	US-09-822-849A-479
C 6	35.4	14.4	4480	17	US-10-312-334-67
C 7	34.8	14.1	2264	15	US-10-252-157-264
C 8	34	13.8	17213	15	US-10-311-455-1455
C 9	34	13.8	786431	15	US-10-412-277-3
C 10	33.2	13.5	560	13	US-10-027-632-216434
C 11	33.2	13.5	560	16	US-10-027-632-216434
C 12	33.2	13.5	3759	13	US-10-282-122A-15979
C 13	32.8	13.3	338702	13	US-10-087-132-292
C 14	32.6	13.3	435	13	US-10-085-783A-40780

C 15	32.6	13.3	435	16	US-10-242-535A-40780	Sequence 40780, A
C 16	32.2	13.1	15674	15	US-10-311-455-335	Sequence 335, App
C 17	32.2	13.1	15674	15	US-10-240-485-29	Sequence 29, Appl
C 18	32.2	13.1	3673778	15	US-10-312-841-2	Sequence 2, Appl
C 19	32	13.0	480	13	US-10-282-122A-15466	Sequence 15466, A
C 20	32	13.0	941	13	US-10-424-599-100533	Sequence 100533,
C 21	32	13.0	3673778	15	US-10-312-841-1	Sequence 1, Appl
C 22	31.8	12.9	422	10	US-09-918-995-2677	Sequence 2677, Ap
C 23	31.8	12.9	680	13	US-10-027-632-152465	Sequence 152465,
C 24	31.8	12.9	680	16	US-10-027-632-152465	Sequence 152465,
C 25	31.8	12.9	1885	17	US-10-641-643-11162	Sequence 1162, Ap
C 26	31.8	12.9	4346	15	US-10-247-671-49	Sequence 49, Appl
C 27	31.8	12.9	251364	15	US-10-175-523-58	Sequence 58, Appl
C 28	31.8	12.9	251364	15	US-10-175-523-61	Sequence 61, Appl
C 29	31.8	12.9	251364	15	US-10-175-523-79	Sequence 79, Appl
C 30	31.6	12.8	472	10	US-09-918-995-28533	Sequence 28533, A
C 31	31.6	12.8	580	13	US-10-027-632-237125	Sequence 237125,
C 32	31.6	12.8	580	13	US-10-027-632-237125	Sequence 237125,
C 33	31.6	12.8	580	13	US-10-027-632-237125	Sequence 237125,
C 34	31.6	12.8	580	13	US-10-027-632-237125	Sequence 237125,
C 35	31.6	12.8	580	16	US-10-027-632-237126	Sequence 237126,
C 36	31.6	12.8	580	16	US-10-027-632-237126	Sequence 237126,
C 37	31.4	12.8	1017	15	US-10-349-680-136	Sequence 237127,
C 38	31.4	12.8	1215	16	US-10-264-237-600	Sequence 136, App
C 39	31.4	12.8	2939	13	US-10-342-887-301	Sequence 600, App
C 40	31.4	12.8	2939	13	US-10-172-119-301	Sequence 301, App
C 41	31.4	12.8	8467	15	US-10-311-455-81	Sequence 301, App
C 42	31.4	12.8	8758	15	US-10-311-455-1092	Sequence 81, Appl
C 43	31.2	12.7	400	9	US-09-960-352-8625	Sequence 1092, Ap
C 44	31.2	12.7	468	9	US-09-864-761-167	Sequence 8625, Ap
C 45	31.2	12.7	530	9	US-09-604-287A-434	Sequence 167, App

ALIGNMENTS

RESULT 1

US-10-198-846-1685/c
; Sequence 1685, Application US/10198846
; Publication No. US20030099974A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Xu, Yongyao
; APPLICANT: Wang, Youzhen
; APPLICANT: Steinmann, Kathleen
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-049
; CURRENT APPLICATION NUMBER: US/10/198,846
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/306,220
; PRIOR FILING DATE: 2001-07-18
; NUMBER OF SEQ ID NOS: 14084
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1685
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 305, 365, 383, 415, 424, 438, 443, 455, 459, 463, 472, 474,
; LOCATION: 485, 488, 507, 535, 553, 557, 562, 594, 619, 635, 637,
; LOCATION: 646, 659, 663, 674, 678, 694, 714, 718, 725, 726, 728, 741,
; LOCATION: 743, 745, 757, 759, 760, 781, 788, 810, 813, 822, 834
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 838, 858, 859
; OTHER INFORMATION: n = A,T,C or G
US-10-198-846-1685

Query Match

15.6%; Score 38.4; DB 15; Length 861;

QY	62	GATTTTAAGTTTACTCCTACTCTGACCCCAAGTGAATTCCTTCTCCAGTCACAGTGCA	121
Db	3035	AATGTTATATTTTTCATTTTAATCCTTATATTCCTTATTCCTCAATTTACTGTGCT	2976
QY	122	A 122	
Db	2975	A 2975	

```

RESULT 3
US-09-879-536-826
; Sequence 826, Application US/09879536
; Patent No. US20020144298A1
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/879,536
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: US 60/088,801
; PRIOR FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 826
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(507)
; OTHER INFORMATION: n = A,T,C or G
US-09-879-536-826

```

```

1  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
2  FILE REFERENCE: PC006
3  CURRENT APPLICATION NUMBER: US/09/764,891
4  CURRENT FILING DATE: 2001-01-17
5  Prior application data removed - consult PALM or file wrapper
6  NUMBER OF SEQ ID NOS: 10231
7  SOFTWARE: PatentIn Ver. 2.0
8  SEQ ID NO 9844

```

```
; LENGTH: 24387
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-9844

Query Match      14.5%; Score 35.6; DB 10; Length 24387;
Best Local Similarity 62.2%; Pred. No. 7.4;
Matches 56; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 62 GATTTAAAGTTTACTCTACTGCTGACCCAGTGAATTCCTTCAGTCACAGTGTCA 121
    |||||
Db 19467 GATTTAAATATATACCTACCTCAGGACCAAGAAAAAGTTAAGCAAGCAGGGTTCCA 19526
    |||||

QY 122 ACCTTACCCCCCAACTGCAACGAGAGTTT 151
    |||||
Db 19527 AGTGCTCTCCCACTTCAACAGATGT 19556
    |||||

RESULT 5
US-09-822-849A-479/c
; Sequence 479, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fachtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 479
; LENGTH: 2803
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-479

Query Match      14.4%; Score 35.4; DB 9; Length 2803;
Best Local Similarity 63.5%; Pred. No. 2.9;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 2 TCATTAACTTTTCAAGGATACCTTTTATTTTCTTTAAGATTCCTGTTCTTTATACACA 61
    |||||
Db 2102 TCATTTCCTCCAGGATAGATATATATTTCTTTGATATTATATATATATATATA 2043
    |||||

QY 62 GATTTAAAGTTTACTCTACTGCTG 86
    |||||
Db 2042 TATATTATATGTACACACACCTG 2018
    |||||

RESULT 6
US-10-312-354-67/c
; Sequence 67, Application US/10312354
; Publication No. US20040101930A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.; JACKSON, Jennifer L.;
; APPLICANT: TANG, Y. Tom; YUE, Henry;
; APPLICANT: ELLIOTT, Vicki S.; TRIBOULEY, Catherine M.;
; APPLICANT: LEE, Ernestine A.; RAMKUMAR, Jayalaxmi;
; APPLICANT: LAL, Preeti G.; XU, Yuming;
; APPLICANT: WARREN, Bridget A.; HAPALIA, April J. A.;
; APPLICANT: BAUGHN, Mariah R.; AZIMZAI, Yalda;
; APPLICANT: BATRA, Sajeev; BURFORD, Neil;
; APPLICANT: YAO, Monique G.; NGUYEN, Dannel B.;
```

```
; APPLICANT: LU, Dyung Aina M.; CHAWLA, Narinder K.;
; APPLICANT: GANDHI, Ameena R.; AU-YOUNG, Janice;
; APPLICANT: ARVIZU, Chandra S.
; TITLE OF INVENTION: SECRETED PROTEINS
; FILE REFERENCE: PI-0133 USN
; CURRENT APPLICATION NUMBER: US/10/312,354
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: US 01/19862
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 60/212,890
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/213,466
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: US 60/214,601
; PRIOR FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: US 60/222,372
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/231,435
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: US 60/232,889
; PRIOR FILING DATE: 2000-09-15
; NUMBER OF SEQ ID NOS: 88
; SOFTWARE: PERL Program
; SEQ ID NO 67
; LENGTH: 4480
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 1623474CB1
US-10-312-354-67

Query Match      14.4%; Score 35.4; DB 17; Length 4480;
Best Local Similarity 63.5%; Pred. No. 3.7;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 2 TCATTAACTTTTCAAGGATACCTTTTATTTTCTTTAAGATTCCTGTTCTTTATACACA 61
    |||||
Db 3771 TCATTTCCTCCAGGATAGATATATATTTCTTTGATATTATATATATATATA 3712
    |||||

QY 62 GATTTAAAGTTTACTCTACTGCTG 86
    |||||
Db 3711 TATATTATATGTACACACACCTG 3687
    |||||

RESULT 7
US-10-252-157-264
; Sequence 264, Application US/10252157
; Publication No. US20030190640A1
; GENERAL INFORMATION:
; APPLICANT: Paris, Mary
; APPLICANT: Pearson, Cecelia I.
; TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
; FILE REFERENCE: PA-0027-1 US
; CURRENT APPLICATION NUMBER: US/10/252,157
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/295,048
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 501
; SOFTWARE: PERL Program
; SEQ ID NO 264
; LENGTH: 2264
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030190640A1 206310.2
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 481-537, 1472, 1846-1848
; OTHER INFORMATION: a, t, c, g, or other
US-10-252-157-264
```


APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 216434
LENGTH: 560
TYPE: DNA
ORGANISM: Human
US-10-027-632-216434

Query Match 13.5%; Score 33.2; DB 16; Length 560;
Best Local Similarity 67.1%; Pred. No. 6.3; Indels 0; Gaps 0;
Matches 47; Conservative 0; Mismatches 23;
QY 7 AACCTTGCAGGATACCTTTATTTCTTTAAAGATTCCTGTGTTTATACACAGATTT 66
Db 389 AAACCTTTCTTCATATACATCTTTATTTCTTTATATATATATATATATTTT 330

QY 67 TAAGTTTACT 76
Db 329 CTAGTTTACT 320

RESULT 12
US-10-282-122A-15979/c
Sequence 15979, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITFA 034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 15979
LENGTH: 3759
TYPE: DNA
ORGANISM: Clostridium acetobutylicum
US-10-282-122A-15979

Query Match 13.5%; Score 33.2; DB 13; Length 3759;
Best Local Similarity 54.0%; Pred. No. 16; Indels 0; Gaps 0;
Matches 68; Conservative 0; Mismatches 58;
QY 20 ATACCTTTTATTTCTTTAAAGATTCCTGTGTTTATACACAGATTTTAAAGTTTACTCT 79
Db 3014 ATTCTTTTAGTCTTTTATAAAATTCGTTAGTTTTCATTATGTTTCTTCTCTCT 2955
QY 80 ACTGCTGACCCAGTGAATTCCTTCCAGTCACAGTGTCAACCTTACCCCCCACTG 139
Db 2954 AATAATACTGATTAACAACTTAAACATCATTTCTTCTTAAACATTTACTCGCCTA 2895
QY 140 CAACGA 145
Db 2894 GAAGGA 2889

RESULT 13
US-10-087-192-292/c
Sequence 292, Application US/10087192
Publication No. US20020182586A1
GENERAL INFORMATION:
APPLICANT: Morris, David W.
APPLICANT: Engelhard, Eric K.
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
TITLE OF INVENTION: CANCER
FILE REFERENCE: 529452000122
CURRENT APPLICATION NUMBER: US/10/087,192
CURRENT FILING DATE: 2002-03-01
PRIOR APPLICATION NUMBER: US 09/747,377
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US 09/798,586
PRIOR FILING DATE: 2001-03-02
NUMBER OF SEQ ID NOS: 2059
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 292
LENGTH: 338702
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(338702)
OTHER INFORMATION: n = A,T,C or G
US-10-087-192-292

Query Match 13.3%; Score 32.8; DB 13; Length 338702;
Best Local Similarity 50.6%; Pred. No. 2.1e+02;
Matches 79; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
QY 33 TTCTTTAAGATTCCTGTGTTTATACACAGATTTTAAAGTTTACTCTTACTGCTGACCCAA 92
Db 212965 TGCCTTACCATTTTATGTTGCGCTTGATCCACTTTAAATATAGGTTTAACTTCTCAGAA 212906
QY 93 GTGAAATTCCTTCTCCAGTGCACAGTGTCAACCTTACCCCACTGCAACGAGAGTTTT 152

```
Db 212905 TCAGGGCTCAGTCACCCCTTGACAGTTTCAAAATCCATTCGCCACCTGAATGGCTCCCACT 212846
QY 153 GAGGGGATCAATACACACCGAGAGTACACAGCCCT 188
Db 212845 GGTGGCAGGAATAAATACTGAAGAGACCCCAAGTCCT 212810

RESULT 14
US-10-085-783A-40780/c
; Sequence 40780, Application US/10085783A
; Publication No. US20040037841A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2002
; CURRENT APPLICATION NUMBER: US/10/085,783A
; CURRENT FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40780
; LENGTH: 435
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (24)..(24)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (27)..(27)
; OTHER INFORMATION: n is a, c, g, or t
US-10-085-783A-40780

Query Match 13.3%; Score 32.6; DB 13; Length 435;
Best Local Similarity 61.7%; Pred. No. 8.6;
Matches 50; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 11 TTTCGAAGGATACCTTTTATTTTCTTTAAAGATTCCTGTTTATACACAGATTTTAAG 70
Db 93 TTGTATGAAATACATTTATTTCTTTAAAGTTACCATTCATATATATCCCAATTATA 34
QY 71 TTACTCTCTACTGCTGACCCA 91
Db 33 TTCACTNCTNCTGTTGATACA 13

US-10-085-783A-40780

Query Match 13.3%; Score 32.6; DB 13; Length 435;
Best Local Similarity 61.7%; Pred. No. 8.6;
Matches 50; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 11 TTTCGAAGGATACCTTTTATTTTCTTTAAAGATTCCTGTTTATACACAGATTTTAAG 70
Db 93 TTGTATGAAATACATTTATTTCTTTAAAGTTACCATTCATATATATCCCAATTATA 34
QY 71 TTACTCTCTACTGCTGACCCA 91
Db 33 TTCACTNCTNCTGTTGATACA 13

US-10-085-783A-40780/c
; Sequence 40780, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
```

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40780
; LENGTH: 435
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (24)..(24)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (27)..(27)
; OTHER INFORMATION: n is a, c, g, or t
US-10-242-535A-40780

Query Match 13.3%; Score 32.6; DB 16; Length 435;
Best Local Similarity 61.7%; Pred. No. 8.6;
Matches 50; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 11 TTTCGAAGGATACCTTTTATTTTCTTTAAAGATTCCTGTTTATACACAGATTTTAAG 70
Db 93 TTGTATGAAATACATTTATTTCTTTAAAGTTACCATTCATATATATCCCAATTATA 34
QY 71 TTACTCTCTACTGCTGACCCA 91
Db 33 TTCACTNCTNCTGTTGATACA 13

Search completed: June 10, 2004, 17:05:34
Job time : 223 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 15:30:58 ; Search time 2535 Seconds
(without alignments)
2897.868 Million cell updates/sec

Title: US-09-674-593-10

Perfect score: 246

Sequence: 1 gtcatcaaccttgcaagga.....ttcatatcaaacccacacata 246

Scoring table:

IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estmus.*

4: em_estpln.*

5: em_estov.*

6: em_estro.*

7: em_estro.*

8: em_estro.*

9: gb_est1.*

10: gb_est2.*

11: gb_est3.*

12: gb_est4.*

13: gb_est5.*

14: gb_est6.*

15: em_estfun.*

16: em_estom.*

17: em_gss_hum.*

18: em_gss_inv.*

19: em_gss_pln.*

20: em_gss_vrt.*

21: em_gss_fun.*

22: em_gss_mus.*

23: em_gss_mus.*

24: em_gss_pro.*

25: em_gss_pro.*

26: em_gss_prg.*

27: em_gss_vrt.*

28: gb_ges1.*

29: gb_ges2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	246	100.0	246	9	AA863443
c 2	244.4	99.3	486	12	BM833169 K-EST0107
3	242	98.4	608	13	BU689416 UI-CF-EC1
c 4	212.4	86.3	845	13	BQ948660 AGENCOURT

RESULT 1

AA863443

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AA863443 246 bp mRNA linear EST 13-MAY-1998
oh05e10.s1 NCI_CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1456938 3',
mRNA sequence.

AA863443.1 GI:2955922

EST.

Homo sapiens (human)

Homo sapiens

Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 246)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/ILNL at:

www-bio.lnl.gov/bbrp/image/image.html

ALIGNMENTS

c 5	174.6	71.0	898	13	BQ650550
c 6	99.4	40.4	963	13	BQ650598
c 7	40.2	16.3	502	14	CB049560
c 8	38	15.4	477	12	BI818248
c 9	37.6	15.3	834	28	BH280049
c 10	37.4	15.2	415	9	AI797543
c 11	37.4	15.2	433	9	AA121152
c 12	37.2	15.1	520	10	BG033754
c 13	37.2	15.1	989	13	BM361014
c 14	37	15.0	1201	29	CNS001KH
c 15	36.8	15.0	669	13	BM315557
c 16	36.6	14.9	860	28	AZ540006
c 17	36.6	14.9	894	28	BH132363
c 18	36.6	14.9	971	28	AZ687607
c 19	36.4	14.8	671	29	CE637367
c 20	36.2	14.7	945	28	BH146881
c 21	36.2	14.7	1201	13	BX375716
c 22	36	14.6	859	28	BZ154702
c 23	36	14.6	976	28	CC293061
c 24	35.8	14.6	595	14	CB168554
c 25	35.6	14.5	479	14	CD446415
c 26	35.6	14.5	1201	9	AL541884
c 27	35.4	14.4	326	12	BM984485
c 28	35.4	14.4	744	9	AA056168
c 29	35.4	14.4	928	12	BG283792
c 30	35.4	14.4	1201	9	AL553722
c 31	35	14.2	846	10	BF698245
c 32	35	14.2	1201	9	AL513871
c 33	34.8	14.1	319	10	AW189673
c 34	34.8	14.1	514	28	AQ825584
c 35	34.8	14.1	556	29	FR0041818
c 36	34.8	14.1	635	14	CB449540
c 37	34.8	14.1	652	13	BQ389577
c 38	34.8	14.1	966	9	AL543767
c 39	34.8	14.1	1201	9	AL551035
c 40	34.6	14.1	422	12	BG508540
c 41	34.6	14.1	576	12	BM521549
c 42	34.6	14.1	584	12	BM521576
c 43	34.6	14.1	632	29	CE727195
c 44	34.6	14.1	1101	29	CNS000FFA
c 45	34.4	14.0	495	12	BM074816

Insert Length: 1358 Std Error: 0.00
Seq primer: -40m13 fwd. RT from Amersham
High quality sequence stop: 208.

FEATURES

source
1. .246
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1456938"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

ORIGIN

Query Match 100.0%; Score 246; DB 9; Length 246;
Best Local Similarity 100.0%; Pred. No. 2.9e-49;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCATTAACTTTGCAAGGATACCTTTTATTTCTTTAAGATTCCTGTTGTTTATACAC 60
Db 1 GTCATTAACTTTGCAAGGATACCTTTTATTTCTTTAAGATTCCTGTTGTTTATACAC 60
QY 61 AGATTTTAAAGTTTACTCTCTACTGCTGACCAAGTGAATTCCTTCCAGTCACAGTGTC 120
Db 61 AGATTTTAAAGTTTACTCTCTACTGCTGACCAAGTGAATTCCTTCCAGTCACAGTGTC 120
QY 121 AACCTCTACCCCACTGCAACGAGAGTTTGGGGGCATCAATCACACCGAGAGTCA 180
Db 121 AACCTCTACCCCACTGCAACGAGAGTTTGGGGGCATCAATCACACCGAGAGTCA 180
QY 181 CAGCCCTCAACACCTGAGTGTGGGGGTAGGATCTGCATTTCTCATATCAACCCC 240
Db 181 CAGCCCTCAACACCTGAGTGTGGGGGTAGGATCTGCATTTCTCATATCAACCCC 240
QY 241 ACACCTA 246
Db 241 ACACCTA 246

RESULT 2

BM833169/c
LOCUS K-85T010772 S5SNU484s1 Homo sapiens cDNA clone S5SNU484s1-12-C04
DEFINITION 486 bp mRNA linear EST 06-MAR-2002
5', mRNA sequence.
ACCESSION BM833169
VERSION BM833169.1 GI:19189578
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 486)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.
21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 12 row: C column: 04

High quality sequence stop: 486.

FEATURES

source
1. .486
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="S5SNU484s1-12-C04"
/sex="M"
/tissue_type="Stomach"
/cell_type="Epithelial"
/lab_host="Top10F"
/clone_lib="S5SNU484s1"

/note="Organ: Stomach; Vector: pTZ18RP1; Site_1: EcoRI; Site_2: NotI; The poly (A)+ RNA was decapped with tobacco acid pyrophosphatase (TAP) and ligated with DNA-RNA linker including EcoRI site by treatment of T4 RNA ligase. The first strand cDNA was synthesized from oligo dT-selected mRNA by priming with dT-tailed vector. The dT-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F' by electroporation method. After analyzing and sequencing about 2,000 ~ 3,000 colonies in original cDNA library, the abundant cDNAs were selected and amplified by PCR reaction using vector region primer including T7 promoter as 5' primer and N(dT)14 as 3' primer. The PCR products were used as template for synthesis of biotinylated single stranded RNA by in vitro transcription reaction. The synthesized RNA probes were hybridized with antisense single stranded cDNAs prepared from original library and incubated with avidin-gel. After removing DNA-RNA hybrids by centrifuge, the subtracted cDNA libraries were constructed by transformation of the remaining DNA into competent cells E. coli Top10F' with electroporation method."

ORIGIN

Query Match 99.3%; Score 244.4; DB 12; Length 486;
Best Local Similarity 99.6%; Pred. No. 6.7e-49;
Matches 245; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTCATTAACTTTGCAAGGATACCTTTTATTTCTTTAAGATTCCTGTTGTTTATACAC 60
Db 466 GTCATTAACTTTGCAAGGATACCTTTTATTTCTTTAAGATTCCTGTTGTTTATACAC 407
QY 61 AGATTTTAAAGTTTACTCTCTACTGCTGACCCCAAGTGAATTCCTTCCAGTCACAGTGTC 120
Db 406 AGATTTTAAAGTTTACTCTCTACTGCTGACCCCAAGTGAATTCCTTCCAGTCACAGTGTC 347
QY 121 AACCTCTACCCCACTGCAACGAGAGTTTGGGGGCATCAATCACACCGAGAGTCA 180
Db 346 AACCTCTACCCCACTGCAACGAGAGTTTGGGGGCATCAATCACACCGAGAGTCA 287
QY 181 CAGCCCTCAACACCTGAGTGTGGGGGTAGGATCTGCATTTCTCATATCAACCCC 240
Db 286 CAGCCCTCAACACCTGAGTGTGGGGGTAGGATCTGCATTTCTCATATCAACCCC 227
QY 241 ACACCTA 246
Db 226 ACACCTA 221

RESULT 3

BU689416
LOCUS BU689416
DEFINITION UI-CF-ECI-adw-1-20-0-UI.s1 UI-CF-ECI Homo sapiens cDNA clone
ACCESSION BU689416
VERSION BU689416.1 GI:23547148
KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 608)
 REFERENCE Bonaldo,M.F., Lennon,G. and Soares,M.B.
 AUTHORS Normalization and subtraction: two approaches to facilitate gene
 TITLE discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8895548
 COMMENT Contact: McCray, PB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul-mccray@uiowa.edu
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com) or from Open Biosystems
 (www.openbiosystems.com).
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES Location/Qualifiers
 source 1..608
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-CF-EC1-adw-1-20-0-UI"
 /tissue_type="Lung"
 /dev_stage="Adult and Fetal"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-CF-EC1"
 /notes="Organ: Lung; Vector: p77T3-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-CF-EC1 is a normalized cDNA library containing the
 following tissue(s): Normal lung from adult and from fetal
 day 64, day 87, week 19 and week 42. The library was
 constructed according to Bonaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into p77T3-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 AAGTGGCTTAC.
 TAG_TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
 and 380-383
 TAG_LIB=UI-CF-EC1
 TAG_SEQ=AAGTGGCTTAC"

ORIGIN
 Query Match 98.4%; Score 242; DB 13; Length 608;
 Best Local Similarity 100.0%; Pred. No. 2.5e-48;
 Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 TAG_SEQ=AAGTGGCTTAC"

QY 5 TTAACCTTTCAGGATACCTTTTATTTCTTAAGATTCCTGTTTATACACAGAT 64
 Db 17 TTAACCTTTCAGGATACCTTTTATTTCTTAAGATTCCTGTTTATACACAGAT 76
 QY 65 TTTAAGTTTACTCCTACTGTCACCCAGTGAATTCCTTCTCCAGTCACAGTGTCAACC 124
 Db 77 TTTAAGTTTACTCCTACTGTCACCCAGTGAATTCCTTCTCCAGTCACAGTGTCAACC 136
 QY 125 TCTACCCCCCACTGCACAGAGTTTGTAGGGGCGATCAATCAACCGAGAAGTCAAGC 184
 Db 137 TCTACCCCCCACTGCACAGAGTTTGTAGGGGCGATCAATCAACCGAGAAGTCAAGC 196

QY 185 CCCTCAACCACTGAGTCTGGGGGTAGGATCTGCAATTTCTTCAATCAACCCACAC 244
 Db 197 CCCTCAACCACTGAGTCTGGGGGTAGGATCTGCAATTTCTTCAATCAACCCACAC 256
 QY 245 TA 246
 Db 257 TA 258

RESULT 4
 BQ948660/c
 LOCUS
 DEFINITION AGNCOURT_8784237 NIH_MGC_43 Homo sapiens cDNA clone IMAGE:6376418
 5' mRNA sequence.
 ACCSSION BQ948660
 VERSION BQ948660.1 GI:22364138
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 845)
 REFERENCE NIH-MGC http://mgi.nci.nih.gov/.
 AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
 TITLE Unpublished (1999)
 JOURNAL
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-x@mail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LICM2558 row: j column: 03
 High quality sequence stop: 534.

FEATURES Location/Qualifiers
 source 1..845
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:6376418"
 /tissue_type="normal pigmented retinal epithelium"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_43"
 /notes="Organ: eye; Vector: pOTB7; Site 1: XhoI; Site 2:
 EcoRI; cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5',
 adaptor: GGCACGAG(C). Library constructed by Ling Hong
 in the laboratory of Gerald W. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC library."

ORIGIN
 Query Match 86.3%; Score 212.4; DB 13; Length 845;
 Best Local Similarity 96.2%; Pred. No. 3.5e-41;
 Matches 227; Conservative 0; Mismatches 8; Indels 1; Gaps 1;
 TAG_SEQ=AAGTGGCTTAC"

QY 11 TTTCAGGATACCTTTTATTTCTTAAGATTCCTGTTTATACACAGATTTAAG 70
 Db 688 TTTCAGGATACCTTTTATTTCTTAAGATTCCTGTTTATACACAGATTTAAG 630
 QY 71 TTTCAGGATACCTTTTATTTCTTAAGATTCCTGTTTATACACAGATTTAAG 130
 Db 629 TTTCAGGATACCTTTTATTTCTTAAGATTCCTGTTTATACACAGATTTAAG 570
 QY 131 CCCCACTGACAGAGATTTTGTAGGGGCGATCAATCAACCGAGAAGTCAAGCCCTCA 190
 Db 569 CCCCACTGACAGAGATTTTGTAGGGGCGATCAATCAACCGAGAAGTCAAGCCCTCA 510
 QY 191 ACCACTGAGGTGTGGGGGTAGGATCTGCATTTCTTCAATCAACCCACACTA 246

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0200

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg Ph D

Email: szhao@tigr.org
 Clones are derived from the rat BAC library CHORI-230
 (http://www.chori.org/bacpac/rat230.htm). For BAC library
 availability, please contact Pieter de Jong (pjejong@mail.cho.org).
 Clones may be purchased from BACPAC Resources
 (http://www.chori.org/bacpac/or ering_information.htm). BAC end
 page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
 Plate: 117 row: E column: 16
 Seq primer: SP6
 Class: BAC ends.

FEATURES

Location/Qualifiers
 1..834

/organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /strain="BN/SSHsd/MCW"
 /db_xref="taxon:10116"
 /clone="CH230-117E16"
 /sex="female"
 /cell_type="Brain"
 /clone_lib="CHORI-230 Segment 1"
 /note="vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI;
 CHORI-230 Rat (BN/SSHsd/MCW) BAC library produced by
 Pieter de Jong"

ORIGIN

Query Match 15.3%; Score 37.6; DB 28; Length 834;
 Best Local Similarity 48.6%; Pred. No. 60;
 Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 9 CCTTTGCAAGATACCTTTTATTTCTTTTAAAGATCTCTGTTTATACACAGATTTA 68
 Db CTTGTCTGCTAAGAGTGTGTTGATTTGTTTTTTCAGTCAACTAAAAAGCTTCTGC 308

QY 69 AGTTTACTCTCTACGCTGACCAAGTCAAAATCTCTTCCAGTCACAGTGCACACTCTA 128
 Db CATTCCTGCCAATCTTAAACCTAAGTGTCTCTCTCAATAATGGTCTTAAGCTG 248

QY 129 CCCCCCACTGCAAGAGAGTTTGTAGGGGCATCAATCACACCGAGAGTCACAGCCCT 188
 Db TTTCTTGAGTCCAGATAGACTTTATATGAGAACCATAGATTTTAAAGCCCTCGCT 188

QY 189 CAACCACTGAGTGTGGGGGTAGGGATCTG 220
 Db GTACAAGTGGGGCGAGGTAAACAGTAGCTG 156

RESULT 10

AI797543/c
 LOCUS AI797543 415 bp mRNA linear EST 18-DEC-1999
 DEFINITION w54h01.xl NCI-CCAP Co3 Homo sapiens cDNA clone IMAGE:2344945 3',
 similar to contains MER22.b2 MER22 repetitive element ; mRNA
 sequence.
 ACCESSION AI797543
 VERSION AI797543.1 GI:5363015
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 415)
 NCIT-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 Contact: Robert Strausberg, Ph.D.
 Email: ccgaps-remail.nih.gov
 Tissue Procurement: Elias Campo, M.D., Michael R. Emmert-Buck,
 M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arraying: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html
 Insert Length: 1718 Std Error: 0.00
 Seg primer: -40Up from Gibco
 High quality sequence stop: 414.
 Location/Qualifiers
 1..415

FEATURES

source
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2344945"
 /sex="pooled"
 /tissue_type="colon"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP Co3"
 /note="vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
 was prepared from 12 pooled bulk tumor samples and primed
 with a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pT7T3 vector. Library went through one round of
 normalization."

ORIGIN

Query Match 15.2%; Score 37.4; DB 9; Length 415;
 Best Local Similarity 55.9%; Pred. No. 71;
 Matches 71; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 25 TTTTATTTTCTTTAAGATTCCTGTTTATACACAGATTTTAAAGTTTACTCTACTGC 84
 Db TTTCTGTTTCGTTTTTTTCTTTTCTATATATATTTTGTGAATCTATTT 191

QY 85 TGACCCCAAGTGAATTCCTTCTCCAGTCACAGTGTCAACCTCTACCCCAACTGCAACG 144
 Db ATTTTAAATCTCTCTCTCTCTCCAGACACATGGCACTGCTATCTCCGAATGGTGTG 131

QY 145 AGAGTTT 151
 Db ATCGTCT 124

RESULT 11

AA121152/c
 LOCUS AA121152 433 bp mRNA linear EST 19-MAY-1997
 DEFINITION z188e01.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
 IMAGE:511704 3', mRNA sequence.
 AA121152
 AA121152.1 GI:1678695
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 433)
 REFERENCE Hillier,L., Lennon,G., Becker,M., Bonaldo,M.P., Chiapelli,B.,
 Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
 Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
 Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
 Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
 Trevaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
 and Marra,M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478
 8889549
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.

TITLE

JOURNAL
 MEDLINE
 PUBMED
 COMMENT

Insert Length: 747 Std Error: 0.00
Seq primer: -40M13 fwd. from Amerham
High quality sequence stop: 371.

FEATURES

source
1..433
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="GDB:3844199"
/db_xref="taxon:9606"
/clone="IMAGE:511704"
/tissue_type="tumor"
/cell_line="T84 carcinoma cell line"
/lab_host="SOLR cells (kanamycin resistant)"
/clone_lib="stratagene colon (#937204)"
/note="Organ: colon; Vector: pBluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. T-84 colonic epithelial cell line. Average
insert size: 1.0 Kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTT 3'"

ORIGIN

Query Match 15.2%; Score 37.4; DB 9; Length 433;
Best Local Similarity 55.9%; Pred. No. 70;
Matches 71; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

QY 25 TTTTATTTCCTTAAGATTCCTGTTGTTTATACACAGATTTTAAGTTTACTCCTACTGC 84
|||||
Db 244 TTTCTGTTCTGTTTTTTTCTTTTCTATATATATTTTGTGAAATCTATTTT 185
|||||
QY 85 TGACCAAGTGAATTCCTTCCAGTCACAGTCAACCTTACCCCAACTGCAAG 144
|||||
Db 184 ATTTTAAATCTCTCTCTCTCCCTCCACACAATGGCACTGCTTATCTCGAATGTTGTG 125
|||||
QY 145 AGAGTTT 151
|||||
Db 124 ATCGTCT 118
|||||

RESULT 12
BG033754/c
LOCUS
DEFINITION
602302013F1 NIH_MGC_87 Homo sapiens cDNA clone IMAGE:4403426 5',
mRNA sequence.
BG033754.1 GI:12426208
BG033754
EST.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 520)
NIH-MGC <http://mgi.nci.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: DCTD/DP
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM10113 row: j column: 03
High quality sequence stop: 223.

FEATURES

source
1..520
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4403426"
/tissue_type="mammary adenocarcinoma, cell line"

/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_87"
/note="Organ: breast; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.383 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

ORIGIN

Query Match 15.1%; Score 37.2; DB 10; Length 520;
Best Local Similarity 53.4%; Pred. No. 78;
Matches 78; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY 25 TTTTATTTCCTTAAGATTCCTGTTGTTTATACACAGATTTTAAGTTTACTCCTACTGC 84
|||||
Db 192 TTTTATTTTTTTTTTTTTTTTGTGATATACAACTACTTCATTTTATTGACTTT 133
|||||
QY 85 TGACCAAGTGAATTCCTTCTCCAGTCACAGTCAACCTTACCCCAACTGCAAG 144
|||||
Db 132 ATATAATAGTGAATCCCTTTAAGCAACCTAGGTATACAGTTGGTGTCCAACCTTCAGG 73
|||||
QY 145 AGAGTTTTCAGGGGATCAATCAGAC 170
|||||
Db 72 CGATTTTCGGGGGAGTAGAACAC 47
|||||

RESULT 13
BX361014/c
LOCUS
DEFINITION
BX361014 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSODI078YP09 5-PRIME, mRNA sequence.
BX361014
EST.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 989)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: secref@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 10967.f For
more information about this cluster, see
[http://www.genoscope.cns.fr/](http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSODI078CH05QP1&cluster=10967.f)
Feng Liang Email: fliang@lifetech.com URL:
<http://fulllength.invitrogen.com/> Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID: CSODI078CH05QPI.

FEATURES

Location/Qualifiers
1..989
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI078YP09"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 15.1%; Score 37.2; DB 13; Length 989;
Best Local Similarity 30.1%; Pred. No. 74;
Matches 68; Conservative 37; Mismatches 121; Indels 0; Gaps 0;

QY 3 CATTAACCTTTGCAAGGATACCTTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAG 62

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 15:30:58 ; Search time 1706 Seconds
(without alignments)
6249.933 Million cell updates/sec

Title: US-09-674-593-10
Perfect score: 246
Sequence: 1 gtcattaaccttgcaagga.....ttcatatcaacccacacta 246

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 segs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:**

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_hcg_hum.*

31: em_hcg_inv.*

32: em_hcg_other.*

33: em_hcg_mus.*

34: em_hcg_pln.*

35: em_hcg_rod.*

36: em_hcg_mam.*

37: em_hcg_vrt.*

38: em_sy.*

39: em_higo_hum.*

40: em_higo_mus.*

41: em_higo_other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	246	100.0	246	6	BD206172	Tumor-ass
2	244.4	99.3	1382	6	BD206166	Tumor-ass
3	244.4	99.3	1382	9	AF181722	Homo sapi
4	244.4	99.3	4377	9	AF181720	Homo sapi
5	244.4	99.3	120029	2	HSJ2828H10	Homo sapi
6	244.4	99.3	152966	9	AL359713	Human DNA
7	39.8	16.2	211530	2	AC106070	Rattus no
8	38.6	15.7	66522	2	AC122565	Mus muscu
9	38.6	15.7	176244	10	AL129208	Mus muscu
10	37.8	15.4	100902	5	AL929176	Zebrafish
11	37.6	15.3	200985	10	AC121787	AC129176
12	37.6	15.3	242754	2	AC129445	Rattus no
13	37.6	15.3	247480	2	AC105857	Rattus no
14	37.4	15.2	212404	10	AC090127	AC090127
15	37.2	15.1	160699	2	AC138731	Pongo pyg
16	37	15.0	223974	2	AC106343	Rattus no
17	37	15.0	227302	2	AC106442	Rattus no
18	36.8	15.0	240977	10	AL513352	Mouse DNA
19	36.6	14.9	103179	10	EX537301	Mouse DNA
20	36.6	14.9	220936	2	AC020256	AC020256
21	36.2	14.7	216849	10	AC129333	AC129333
22	36.2	14.7	243290	10	AL663088	Mouse DNA
23	36	14.6	110000	2	AC095863_05	Continuation (6 of
24	36	14.6	120330	8	AP004361	Oryza sat
25	36	14.6	141293	2	AP004337	Oryza sat
26	36	14.6	142446	8	AP005126	Oryza sat
27	36	14.6	147750	10	AC122503	AC122503
28	36	14.6	180230	10	AL124692	Mouse muscu
29	36	14.6	185066	10	AL844560	Mouse DNA
30	36	14.6	218951	2	AC111418	AC111418
31	36	14.6	259077	2	AC105877	Rattus no
32	35.8	14.6	87065	9	HSJ1174H9	AC105877
33	35.8	14.6	135389	9	AL512292	Human DNA
34	35.8	14.6	181016	9	AC068273	AC068273
35	35.8	14.6	254160	2	AC060761	Homo sapi
36	35.6	14.5	507	6	BD210936	Mouse muscu
37	35.6	14.5	43661	9	AC004493	Human gen
38	35.6	14.5	132029	9	HS795623	AC004493
39	35.6	14.5	148724	2	AC010778	Homo sapi
40	35.6	14.5	155017	2	AC068583	AC010778
41	35.6	14.5	186278	9	AC079176	AC068583
42	35.6	14.5	204521	2	AC125005	Homo sapi
43	35.6	14.5	259508	2	AC112345	AC125005
44	35.4	14.4	630	11	EV004534	Rattus no
45	35.4	14.4	1800	9	HSM801200	BV004534 S208P6386

ALIGNMENTS

RESULT 1
BD206172
LOCUS
DEFINITION
Tumor-associated antigen encoded by reverse strand of novel gene
expressed unevenly.
ACCESSION
BD206172
VERSION
JP 2002514400-A/7
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 246)
AUTHORS
Eynde,B.V.D. and Falleur,T.B.
TITLE
Tumor-associated antigen encoded by reverse strand of novel gene

Pred. No. is the number of results predicted by chance to have a

expressed unevenly
 Patent: JP 2002514400-A 7 21-MAY-2002;
 LUDWIG INSTITUTE FOR CANCER RESEARCH
 OS Homo sapiens (human)
 PN JP 2002514400-A/7
 PD 21-MAY-2002
 PF 13-MAY-1999 JP 2000548350
 PR 13-MAY-1998 US 60/085318
 PI BENOIT VAN DEN EYNDE, THIERRY BOON FALLEUR
 PC C12N15/09, A61K38/00, A61P35/00, C07K14/47, C07K16/18,
 PC C12N1/15,
 PC C12N1/19, C12N1/21, C12N5/10, C12P21/02, C12P21/08, C12Q1/68// PC
 A61K35/12,
 PC C12N15/00, A61K37/02, C12N5/00
 CC Tumor-associated antigen encoded by reverse strand of novel
 CC unevenly.
 CC gene expressed
 FH Key Location/Qualifiers
 FT source 1..246
 FT source /organism="Homo sapiens (human)"
 FT source /mol_type="genomic DNA"
 FT source /db_xref="taxon:9606"

FEATURES
 source
 Location/Qualifiers
 1..246
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 100.0%; Score 246; DB 6; Length 246;
 Best Local Similarity 100.0%; Pred. No. 1.7e-61;
 Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCATTACCTTGCAGGATACCTTTTATTTCTTTAAAGATTCTCTGTTGTTTATAC 60
 Db 1 GTCATTACCTTGCAGGATACCTTTTATTTCTTTAAAGATTCTCTGTTGTTTATAC 60
 QY 61 AGATTTTAAAGTTTACTCTCTGCTGACCCAAAGTGAATTCCTTCCAGTCACAGTGTG 120
 Db 61 AGATTTTAAAGTTTACTCTCTGCTGACCCAAAGTGAATTCCTTCCAGTCACAGTGTG 120
 QY 121 AACCTCTACCCCTCAACCACTGCAACGAGAGTTTGGGGGATCAATCAACCGAGAAGTCA 180
 Db 121 AACCTCTACCCCTCAACCACTGCAACGAGAGTTTGGGGGATCAATCAACCGAGAAGTCA 180
 QY 181 CAGCCCTCAACCACTGAGGTGTGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC 240
 Db 181 CAGCCCTCAACCACTGAGGTGTGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC 240
 QY 241 ACACCTA 246
 Db 241 ACACCTA 246

RESULT 2
 BD206166 1382 bp DNA linear PAT 17-JUL-2003
 LOCUS Tumor-associated antigen encoded by reverse strand of novel gene
 DEFINITION expressed unevenly.
 ACCESSION BD206166
 VERSION BD206166.1 GI:33015936
 KEYWORDS JP 2002514400-A/1.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1382)
 AUTHORS Eynde, B.V.D. and Falleur, T.B.
 TITLE Tumor-associated antigen encoded by reverse strand of novel gene
 JOURNAL Patent: JP 2002514400-A 1 21-MAY-2002;
 LUDWIG INSTITUTE FOR CANCER RESEARCH
 OS Homo sapiens (human)
 PN JP 2002514400-A/1
 PD 21-MAY-2002

PF 13-MAY-1999 JP 2000548350
 PI 13-MAY-1998 US 60/085318
 PC C12N15/09, A61K38/00, A61P35/00, C07K14/47, C07K16/18,
 PC C12N1/15,
 PC C12N1/19, C12N1/21, C12N5/10, C12P21/02, C12P21/08, C12Q1/68// PC
 A61K35/12,
 PC C12N15/00, A61K37/02, C12N5/00
 CC Tumor-associated antigen encoded by reverse strand of novel
 CC unevenly.
 CC gene expressed
 FH Key Location/Qualifiers
 FT CDS 1..1382
 FT CDS 738..989.
 FT CDS /organism="Homo sapiens"
 FT CDS /mol_type="genomic DNA"
 FT CDS /db_xref="taxon:9606"

FEATURES
 source
 Location/Qualifiers
 1..1382
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 99.3%; Score 244.4; DB 6; Length 1382;
 Best Local Similarity 99.6%; Pred. No. 5.3e-61;
 Matches 245; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCATTACCTTGCAGGATACCTTTTATTTCTTTAAAGATTCTCTGTTGTTTATAC 60
 Db 268 GTCATTACCTTGCAGGATACCTTTTATTTCTTTAAAGATTCTCTGTTGTTTATAC 327
 QY 61 AGATTTTAAAGTTTACTCTCTGCTGACCCAAAGTGAATTCCTTCCAGTCACAGTGTG 120
 Db 328 AGATTTTAAAGTTTACTCTCTGCTGACCCAAAGTGAATTCCTTCCAGTCACAGTGTG 387
 QY 121 AACCTCTACCCCTCAACCACTGCAACGAGAGTTTGGGGGATCAATCAACCGAGAAGTCA 180
 Db 388 AACCTCTACCCCTCAACCACTGCAACGAGAGTTTGGGGGATCAATCAACCGAGAAGTCA 447
 QY 181 CAGCCCTCAACCACTGAGGTGTGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC 240
 Db 448 CAGCCCTCAACCACTGAGGTGTGGGGGTAGGGATCTGCATTTCTTCATATCAACCCC 507
 QY 241 ACACCTA 246
 Db 508 ACACCTA 513

RESULT 3
 AF181722 1382 bp mRNA linear PRI 10-JAN-2000
 LOCUS Homo sapiens RU2AS (RU2) mRNA, complete cds.
 DEFINITION AF181722
 ACCESSION AF181722.1 GI:6684531
 VERSION AF181722.1
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 1382)
 AUTHORS Van Den Eynde, B.J., Gaugler, B., Probst-Keppler, M., Michaux, L.,
 Devuyt, O., Lorge, P., Weynants, P. and Boon, T.
 TITLE A new antigen recognized by cytolytic T lymphocytes on a human
 kidney tumor results from reverse strand transcription
 JOURNAL J. Exp. Med. 190 (12), 1793-1800 (1999)
 MEDLINE 20069887
 PUBMED 10601354
 REFERENCE 2 (bases 1 to 1382)
 AUTHORS Van den Eynde, B.J., Gaugler, B. and Pilotte, L.
 TITLE Direct Submission
 JOURNAL Submitted (30-AUG-1999) Ludwig Institute for Cancer Research,
 Avenue Hippocrate, 74, Brussels 1200, Belgium
 FEATURES Location/Qualifiers
 source 1..1382
 /organism="Homo sapiens"
 /mol_type="mRNA"

misc_feature /note="match: GSS: Em:B63526"
32. .494
misc_feature /note="match: GSS: Em:AQ315706"
70. .589
repeat_region /note="match: GSS: Em:AQ285677"
179. .362
repeat_region /note="MIR repeat: matches 65. .262 of consensus"
521. .582
repeat_region /note="L2 repeat: matches 2683. .2741 of consensus"
720. .1022
repeat_region /note="AluSx repeat: matches 1. .303 of consensus"
1327. .1448
repeat_region /note="AluJo/FRAM repeat: matches 181. .298 of consensus"
2264. .2575
repeat_region /note="AluSq repeat: matches 1. .311 of consensus"
3014. .3296
misc_feature /note="match: STS: Em:G27290"
3014. .3093
repeat_region /note="40 copies 2 mer ga 75% conserved"
5024. .5477
repeat_region /note="L1M1 repeat: matches 1012. .1598 of consensus"
5581. .5959
repeat_region /note="AluSg repeat: matches 18. .294 of consensus"
5986. .6359
repeat_region /note="L1MEC repeat: matches 2228. .2259 of consensus"
6367. .6515
repeat_region /note="FRAM_C repeat: matches 1. .143 of consensus"
6971. .7198
repeat_region /note="L2 repeat: matches 1512. .1735 of consensus"
8403. .8838
repeat_region /note="L2 repeat: matches 1512. .1735 of consensus"
8799. .9441
misc_feature /note="match: GSS: Em:B54691"
complement(9393. .9853)
misc_feature /note="match: GSS: Em:AQ285817"
10311. .10437
repeat_region /note="L2 repeat: matches 2612. .2747 of consensus"
10710. .11022
repeat_region /note="AluJo repeat: matches 1. .309 of consensus"
12006. .12156
repeat_region /note="L2 repeat: matches 2594. .2748 of consensus"
12182. .13011
repeat_region /note="L2 repeat: matches 1374. .2344 of consensus"
13404. .13631
repeat_region /note="114 copies 2 mer at 55% conserved"
13479. .13630
repeat_region /note="38 copies 4 mer tata 60% conserved"
13785. .14820
repeat_region /note="L1ME repeat: matches 4766. .5798 of consensus"
14928. .16008
repeat_region /note="L1PB3 repeat: matches 5106. .6149 of consensus"
16029. .16110
repeat_region /note="L1PB3 repeat: matches 5054. .5133 of consensus"
16428. .16829
repeat_region /note="L1MC3 repeat: matches 6189. .6603 of consensus"
16868. .17150
repeat_region /note="AluSx repeat: matches 1. .285 of consensus"
17246. .17991
repeat_region /note="L1MC4 repeat: matches 6608. .7343 of consensus"
17992. .18675
repeat_region /note="MER67C repeat: matches 1. .710 of consensus"
18676. .18729
repeat_region /note="L1MC4 repeat: matches 7343. .7396 of consensus"
18731. .19458
repeat_region /note="MER4D repeat: matches 232. .973 of consensus"
complement(18929. .19542)
misc_feature /note="match: GSS: Em:AQ342961"
complement(19031. .19542)
misc_feature /note="match: GSS: Em:AQ053552"
complement(19065. .19495)
misc_feature /note="match: GSS: Em:AQ545892"
complement(19130. .19546)
/note="match: GSS: Em:B94529"

misc_feature 19265. .19495
/note="match: GSS: Em:AQ373351 Em:AQ375803
match: STS: Em:G59266"
complement(19269. .19546)
misc_feature /note="match: GSS: Em:AQ663911"
complement(19325. .19546)
misc_feature /note="match: GSS: Em:AQ021494"
complement(19337. .19501)
misc_feature /note="match: GSS: Em:AQ268095"
complement(19370. .19546)
/note="match: GSS: Em:AQ128461"
19566. .19739
repeat_region /note="AluSg/x repeat: matches 126. .295 of consensus"
19740. .20100
repeat_region /note="L1MD3 repeat: matches 7391. .7739 of consensus"
21121. .21429
repeat_region /note="AluJo repeat: matches 1. .310 of consensus"
22589. .22728
repeat_region /note="70 copies 2 mer aa 60% conserved"
22591. .22670
repeat_region /note="20 copies 4 mer aaag 80% conserved"
22685. .22761
repeat_region /note="19 copies 4 mer aagg 93% conserved"
23098. .23393
repeat_region /note="AluJb repeat: matches 4. .298 of consensus"
23468. .23578
repeat_region /note="L2 repeat: matches 2580. .2700 of consensus"
23980. .24292
repeat_region /note="AluSg repeat: matches 2. .310 of consensus"
complement(27208. .27677)
misc_feature /note="match: GSS: Em:AQ702871"
27819. .27957
repeat_region /note="AluJb repeat: matches 163. .299 of consensus"
28701. .28930
repeat_region /note="MLTIG repeat: matches 32. .301 of consensus"
29080. .29171
repeat_region /note="MLT1H repeat: matches 433. .526 of consensus"
29303. .29711
repeat_region /note="L2 repeat: matches 2323. .2710 of consensus"
29712. .29834
repeat_region /note="MLTIG repeat: matches 29. .147 of consensus"
30140. .30547
repeat_region /note="L2 repeat: matches 1806. .2252 of consensus"
complement(30720. .31207)
misc_feature /note="match: STS: Em:HSJ17T"
30919. .31120
repeat_region /note="MER58A repeat: matches 1. .208 of consensus"
32109. .32472
repeat_region /note="THE1C repeat: matches 1. .371 of consensus"
33360. .33511
repeat_region /note="MIR repeat: matches 98. .250 of consensus"
34031. .34331
repeat_region /note="AluSx repeat: matches 1. .297 of consensus"
34621. .34732
repeat_region /note="MIR repeat: matches 120. .232 of consensus"
34888. .35007
repeat_region /note="L2 repeat: matches 1836. .1959 of consensus"
35164. .35461
repeat_region /note="AluSg repeat: matches 2. .298 of consensus"
35532. .35990
repeat_region /note="L2 repeat: matches 2254. .2710 of consensus"
36631. .36790
repeat_region /note="L2 repeat: matches 2342. .2501 of consensus"
37251. .37544
repeat_region /note="AluSx repeat: matches 1. .294 of consensus"
38446. .38493
repeat_region /note="12 copies 4 mer caca 75% conserved"

Query Match 99.3%; Score 244.4; DB 9; Length 152966;

Best Local Similarity 99.6%; Pred. No. 6.5e-61;

Matches 245; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY

1 GTCATTAACTTTGCAGGATACCTTTTATTTTCTTAAGATTCCTGTTGTTTATACAC 60

```

|||||
53689 GTCACTAACTTGCAGATACCTTTTATTTCTTTAAAGATTCTGTGTTATACAC 53748
|||||
QY 61 AGATTTTAAAGTTTACTTCTACTGCTGACCCAGTGAATTCCTTCCAGTCCACAGTGTC 120
|||||
Db 53749 AGATTTTAAAGTTTACTTCTACTGCTGACCCAGTGAATTCCTTCCAGTCCACAGTGTC 53808
|||||
QY 121 AACCTCTACCCCACTGCAACGAGAGTTTTCAGGGGCATCAATCAACCCGAGAAGTCA 180
|||||
Db 53809 AACCTCTACCCCACTGCAACGAGAGTTTTCAGGGGCATCAATCAACCCGAGAAGTCA 53868
|||||
QY 181 CAGCCCTCAACACATGAGAGTGTGGGGGTAGGATCTGCATTTCTTCATATCAACCCC 240
|||||
Db 53869 CAGCCCTCAACACATGAGAGTGTGGGGGTAGGATCTGCATTTCTTCATATCAACCCC 53928
|||||
QY 241 ACACATA 246
|||||
Db 53929 ACACATA 53934

RESULT 7
AC106070/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-128D5, *** SEQUENCING IN PROGRESS
***, 2 unordered pieces.
AC106070
VERSION AC106070.5 GI:30521767
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 211530)
Muzny,D,Marie, Metzker,M,Lee, Abranzon,S, Adams,C, Alder,J,
Allen,C, Allen,H, Alsbrooks,S, Amin,A, Anguiano,D,
Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benahmed,F,
Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E,
Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A,
Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,
Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Divya,K,
Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K,
Egan,A, Escoto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,
Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,N, Foster,P,
Fraser,C,M, Gabisi,A, Ganta,K, Garcia,A, Garner,T, Garza,M,
Gebregorgis,B, Geer,K, Gill,R, Grady,M, Guerra,W, Guevara,M,
Gunaratne,P, Haaland,W, Hamil,C, Hamilton,C, Hamilton,K,
Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,
Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hognes,M,
Hollins,B, Howells,S, Hulyk,S, Hume,J, Idlebird,D, Jackson,A,
Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolivet,A,
Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C,
Kwais,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,
Lorensuwa,L, Loulaeged,H, Lozado,R,J, Lu,X, Ma,J,
Maheshwari,M, Mahindaratne,M, Mahmoud,M, Malloy,K, Mangum,A,
Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E,
Mawhney,S, McLeod,M,P, McNeill,T,Z, Meenen,E,
Milosavljevic,A, Miner,K, Minja,E, Montemayor,J, Moore,S,
Morgan,M, Morris,K, Morris,S, Munidasa,M, Murphy,M, Nair,L,
Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S,
Nwaokemele,O, Okwuonu,G, Olarnpunsagoon,A, Pal,S, Parks,K,
Pasternak,S, Paul,H, Perez,A, Perez,L, Pfankech,C,
Plopper,F, Poindecker,A, Popovic,D, Primus,E, Pu,L-L,
Puazo,M, Quirroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R,
Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F,
Rives,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J,
Sanders,W, Savary,G, Scherer,S, Scott,G, Shatsman,S, Shen,H,
Shetty,J, Shvartsbeyn,A, Sibson,I, Sitter,C,D, Smajs,D,

```

```

Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willis,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 211530)
Worley,K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 211530)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHGQ
Center clone name: CH230-128D5
----- Summary Statistics
Assembly program: Atlas 3.0;
Consensus quality: 195751 bases at least Q40
Consensus quality: 198846 bases at least Q30
Consensus quality: 200700 bases at least Q20
Estimated insert size: 204266; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
consists of 2 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
* 1 210290: contig of 210290 bp in length
* 210291 210390: gap of unknown length
* 210391 211530: contig of 1140 bp in length.
Location/Qualifiers
1. 211530
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clones="CH230-128D5"
1. 11797
misc_feature
FEATURES
source

```

RESULT 8
 AC122565/c
 LOCUS
 DEFINITION
 AC122565
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus
 Mus musculus (house mouse)
 HTG; HTGS PHASE0
 GI-25229256
 66522 bp
 DNA
 linear
 HTG 24-NOV-2002
 Mus musculus clone RP24-239A21, LOW-PASS SEQUENCE SAMPLING.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 Birren,B., Nusbaum,C. and Lander,E.
 1 (bases 1 to 66522)
 Mus musculus, clone RP24-239A21
 Unpublished
 2 (bases 1 to 66522)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
 Bouckgaier,B., Brown,A., Camarata,J., Campotiano,A., Chang,J.,
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Fero,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
 Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
 Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
 Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,
 Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
 Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neil,D., Oliveri,J., Peterson,K., Phunkhang,P., Pierre,N.,
 Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
 Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,

```

* NOTE: This record contains 83 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1      652: contig of 652 bp in length
*      653      752: gap of 100 bp
*      753      1459: contig of 707 bp in length
*      1460      1559: gap of 100 bp
*      1560      2283: contig of 724 bp in length
*      2284      2383: gap of 100 bp
*      2384      3079: contig of 696 bp in length
*      3080      3179: gap of 100 bp
*      3180      3878: contig of 699 bp in length
*      3879      3978: gap of 100 bp
*      3979      4684: contig of 706 bp in length
*      4685      4784: gap of 100 bp
*      4785      5490: contig of 706 bp in length
*      5491      5590: gap of 100 bp
*      5591      6271: contig of 681 bp in length
*      6272      6371: gap of 100 bp
*      6372      7074: contig of 703 bp in length
*      7075      7174: gap of 100 bp
*      7175      7886: contig of 712 bp in length

```

```
* 7887 7886: gap of 100 bp
* 7987 8705: contig of 719 bp in length
* 8706 8805: gap of 100 bp
* 8806 9511: contig of 706 bp in length
* 9512 9611: gap of 100 bp
* 9612 10321: contig of 710 bp in length
* 10322 10421: gap of 100 bp
* 10422 11116: contig of 695 bp in length
* 11117 11216: gap of 100 bp
* 11217 11917: contig of 701 bp in length
* 11918 12017: gap of 100 bp
* 12018 12726: contig of 709 bp in length
* 12727 12826: gap of 100 bp
* 12827 13529: contig of 703 bp in length
* 13530 13629: gap of 100 bp
* 13630 14336: contig of 707 bp in length
* 14337 14436: gap of 100 bp
* 14437 15153: contig of 717 bp in length
* 15154 15253: gap of 100 bp
* 15254 15953: contig of 700 bp in length
* 15954 16053: gap of 100 bp
* 16054 16747: contig of 694 bp in length
* 16748 16847: gap of 100 bp
* 16848 17572: contig of 725 bp in length
* 17573 17672: gap of 100 bp
* 17673 18382: contig of 710 bp in length
* 18383 18482: gap of 100 bp
* 18483 19190: contig of 708 bp in length
* 19191 19290: gap of 100 bp
* 19291 19979: contig of 689 bp in length
* 19980 20079: gap of 100 bp
* 20080 20777: contig of 698 bp in length
* 20778 20877: gap of 100 bp
* 20879 21581: contig of 704 bp in length
* 21582 21681: gap of 100 bp
* 21682 22377: contig of 696 bp in length
* 22378 22477: gap of 100 bp
* 22479 23183: contig of 706 bp in length
* 23184 23283: gap of 100 bp
* 23284 23986: contig of 703 bp in length
* 23987 24086: gap of 100 bp
* 24087 24993: contig of 707 bp in length
* 24994 24893: gap of 100 bp
* 24894 25592: contig of 699 bp in length
* 25593 25692: gap of 100 bp
* 25693 26417: contig of 725 bp in length
* 26418 26517: gap of 100 bp
* 26519 27331: contig of 714 bp in length
* 27332 27331: gap of 100 bp
* 27332 28048: contig of 717 bp in length
* 28049 28148: gap of 100 bp
* 28149 28843: contig of 695 bp in length
* 28844 28943: gap of 100 bp
* 28944 29638: contig of 695 bp in length
* 29639 29738: gap of 100 bp
* 29739 30456: contig of 718 bp in length
* 30457 30556: gap of 100 bp
* 30557 31234: contig of 678 bp in length
* 31235 31334: gap of 100 bp
* 31335 32037: contig of 703 bp in length
* 32038 32137: gap of 100 bp
* 32138 32846: contig of 709 bp in length
* 32847 32946: gap of 100 bp
* 32947 33664: contig of 718 bp in length
* 33665 33764: gap of 100 bp
* 33765 34478: contig of 714 bp in length
* 34479 34578: gap of 100 bp
* 34579 35275: contig of 697 bp in length
* 35276 35375: gap of 100 bp
* 35376 36084: contig of 709 bp in length
* 36085 36184: gap of 100 bp
* 36185 36896: contig of 712 bp in length
* 36897 36996: gap of 100 bp

* 36997 37686: contig of 690 bp in length
* 37687 37786: gap of 100 bp
* 37787 38472: contig of 686 bp in length
* 38473 38572: gap of 100 bp
* 38573 39274: contig of 702 bp in length
* 39275 39374: gap of 100 bp
* 39375 40071: contig of 697 bp in length
* 40072 40171: gap of 100 bp
* 40172 40869: contig of 698 bp in length
* 40870 40969: gap of 100 bp
* 40970 41674: contig of 705 bp in length
* 41675 41774: gap of 100 bp
* 41775 42453: contig of 679 bp in length
* 42454 42553: gap of 100 bp
* 42554 43261: contig of 708 bp in length
* 43262 43361: gap of 100 bp
* 43362 44073: contig of 712 bp in length
* 44074 44173: gap of 100 bp
* 44174 44868: contig of 695 bp in length
* 44869 44968: gap of 100 bp
* 44969 45662: contig of 694 bp in length
* 45670 45762: gap of 100 bp
* 45763 46455: contig of 693 bp in length

Query Match 15.7%; Score 38,6; DB 2; Length 66522;
Best Local Similarity 52.1%; Pred. No. 2,4;
Matches 86; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 33 TTCCTTAAGATTCCTGTTGTTTATACACAGATTATTAAGTTTACTCTCTACTGTCGACCCAA 92
Db 21504 TTCTTTTCATGTCAGGCTTTCAATACCACTTATTAAATGCAACTCAATGACGACCCCTC 21445

QY 93 GTGAATTCCTTCTCCAGTCACAGTGTCACCTCTACCCCACTGCAACGAGAGTTT 152
Db 21444 TGGCATAAAGTCTCCAGTAACTGCTATGCTTTCACTTTTCAGCAGAGTGAGACCTGA 21385

QY 153 GAGGGGCATCAATCACACCGAAGTCACAGCCCTCAACCACTG 197
Db 21384 GGCAGATATCAGTCAGCCAGGCACTTTGCTCCCTAAACAGCAG 21340

RESULT 9
AC129208 176244 bp DNA linear ROD 27-NOV-2003
LOCUS Mus musculus BAC clone RP24-282E2 from chromosome 15, complete
DEFINITION sequence.
ACCESSION AC129208
VERSION AC129208.4 GI:33285229
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 176244)
AUTHORS Shah,N. and Bielicki,L.
TITLE The sequence of Mus musculus BAC clone RP24-282E2
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 176244)
AUTHORS Wilson,R.
TITLE Sequencing of Mus musculus
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 176244)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (27-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 176244)
AUTHORS McPherson,J.D. and Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (24-APR-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 176244)
AUTHORS Wilson,R.K.
```

TITLE
JOURNAL
REFERENCE
 6 (bases 1 to 176244)
AUTHORS
 Wilson,R.
TITLE
JOURNAL
COMMENT
 Direct Submission
 Submitted (26-JUL-2003) Genome Sequencing Center, 4444 Forest Park
 Parkway, St. Louis, MO 63108, USA
 Submitted (27-NOV-2003) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Jul 26, 2003 this sequence version replaced gi:30089830.
 ----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu>
 Contact: submissions@wustl.edu
 ----- Summary Statistics
 Center project name: M_BB0282E02

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry, or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one subclone; and the assembly was confirmed by
 restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren,
 Department of Genetics, Washington University, St. Louis MO. For
 additional information about the map position of this sequence, see
<http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-24 BAC Library has been constructed by Pieter de Jong and
 coworkers (<http://www.chori.org>) from male C57Bl/6J mouse spleen
 and/or brain genomic DNA. The clone and detailed information can be
 obtained from Pieter de Jong and coworkers at <http://www.chori.org>

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone. This clone is
 overlapped by AC098719.

FEATURES

Source	Location/Qualifiers
1. 176244	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/chromosome="15"
	/map="15"
	/clone="RP24-282E2"
	/clone_lib="RPCI-24"
3420..3647	/rpt_family="MIR"
4355..4594	/rpt_family="ERVK"
5094..5356	/rpt_family="ERVK"
5399..6422	/rpt_family="ERVK"
6416..7025	/rpt_family="ERVK"
11493..11743	/rpt_family="L1"
11743..12500	/rpt_family="L1"
12521..12903	/rpt_family="L1"
14087..14178	/rpt_family="MALR"
	/rpt_family="ERV1"
repeat_region	14603..14907 /rpt_family="ERVK"
repeat_region	17163..17284 /rpt_family="ERVK"
repeat_region	18154..18606 /rpt_family="L1"
repeat_region	18667..18752 /rpt_family="L1"
repeat_region	18845..18955 /rpt_family="Alu"
repeat_region	19680..19899 /rpt_family="B4"
repeat_region	20285..20638 /rpt_family="MALR"
repeat_region	21034..21106 /rpt_family="ID"
repeat_region	21673..21740 /rpt_family="ERV1"
repeat_region	22063..22324 /rpt_family="ERVL"
repeat_region	22514..22980 /rpt_family="L1"
repeat_region	23845..24018 /rpt_family="L2"
repeat_region	25126..25605 /rpt_family="ERVK"
repeat_region	25722..25880 /rpt_family="ERVK"
repeat_region	25856..25961 /rpt_family="ERVK"
repeat_region	25937..26593 /rpt_family="ERVK"
repeat_region	26594..26874 /rpt_family="ERVK"
repeat_region	26871..27626 /rpt_family="ERVK"
repeat_region	27629..28130 /rpt_family="ERVK"
repeat_region	28459..28693 /rpt_family="L2"
repeat_region	29279..29640 /rpt_family="L2"
repeat_region	31661..31840 /rpt_family="L2"
repeat_region	31877..32023 /rpt_family="Alu"
repeat_region	32303..32357 /rpt_family="ERV1"
repeat_region	32399..32896 /rpt_family="L1"
repeat_region	32908..33064 /rpt_family="L1"
repeat_region	33242..33338 /rpt_family="MER1_type"
repeat_region	36084..36137 /rpt_family="B2"
repeat_region	36340..36756 /rpt_family="L1"
repeat_region	36757..37043 /rpt_family="L1"
repeat_region	37054..39354 /rpt_family="L1"
repeat_region	39426..39887 /rpt_family="L1"
repeat_region	39891..40039 /rpt_family="ERVK"
repeat_region	40040..40114 /rpt_family="ERVK"
repeat_region	40328..40975 /rpt_family="ERVK"
repeat_region	42476..43157 /rpt_family="L1"
repeat_region	43153..43264

REFERENCE AUTHORS TITLE JOURNAL	REFERENCE AUTHORS TITLE JOURNAL	COMMENT
--	--	---------

5 (bases 1 to 200985)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (29-DEC-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
6 (bases 1 to 200985)
Wilson,R.
Direct Submission
Submitted (11-NOV-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Dec 29, 2002 this sequence version replaced gi:22475661.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: submissions@watson.wustl.edu
----- Summary Statistics
Center project name: M BA0393K24

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu>

SOURCE INFORMATION:

The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa and Minako Tateno in the laboratory of Pieter de Jong (<http://www.chori.org>) from female C57BL/6J mouse kidney and/or brain genomic DNA. The clone and detailed information can be obtained from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>.

NEIGHBORING SEQUENCE INFORMATION:

This sequence is the entire insert of the clone.

FEATURES source

RRS	source	Location/Qualifiers
		1. 200985
		/organism="Mus musculus"
		/mol_type="genomic DNA"
		/db_xref="taxon:10090"
		/chromosome="9"
		/map="9"
		/clone="RP23-393K24"
		/clone_lib="RPCI-23"
repeat_region		41..171
repeat_region		rpt_family="L1"
repeat_region		841..1023
repeat_region		rpt_family="B2"
repeat_region		2407..2565
repeat_region		rpt_family="B4"
repeat_region		2549..2964
repeat_region		rpt_family="ERVK"
repeat_region		3120..3418
repeat_region		rpt_family="MaLR"
repeat_region		3531..3744
repeat_region		rpt_family="MaLR"
repeat_region		5631..5771
repeat_region		rpt_family="B4"
repeat_region		5794..5920

```

repeat_region 23268..23415
/rpt family="Alu"
repeat_region 23416..23760
/rpt family="MaLR"
repeat_region 23761..24391
/rpt family="L1"
repeat_region 25082..25241
/rpt family="L1"
repeat_region 25235..25804
/rpt family="L1"
repeat_region 25787..25853
/rpt family="L1"
repeat_region 26146..26377
/rpt family="ERVK"
repeat_region 26380..26746
/rpt family="ERVK"
repeat_region 26768..26912
/rpt family="Alu"
repeat_region 27005..27331
/rpt family="L1"
repeat_region 27584..27776
/rpt family="B2"
repeat_region 27805..27875
/rpt family="ID"
repeat_region 27891..28013
/rpt family="B4"

Query Match 15.3% Score 37.6; DB 10; Length 200985;
Best Local Similarity 50.0%; Pred. No. 5;
Matches 94; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 25 TTTTATTTCTTTAAGATTCCTGTGTTTATACACAGATTGTTAGTTTACTCTTACTGTC 84
Db 23825 TTTTCTCTCTGAGGATTTCAATAGCTTCTTTGTTTGTATATAGTGTGTTTAAAC 23766

QY 85 TGACCAAGTGAATTCCTTCCAGTCACAGTGTCAACCTTACCCCCCACTGCAACG 144
Db 23765 TAACATCTTTGTTGAGCTTCTTATGTTATGATAAATACCATGACCAAGTTGCTAAG 23706

QY 145 AGAGTTTGTAGGGGCATCAATCACACCGAAGTCACAGCCCTCAACCACTGAGGTGTG 204
Db 23705 AGAAAAAGGGGGTTTCTTCTTCTGATCACAGTCCATCACTGGAAGGAGTCAG 23646

QY 205 GGGGGGGTA 212
Db 23645 GGCAGAA 23638

RESULT 12
AC129445
LOCUS AC129445
DEFINITION Rattus norvegicus clone CH230-13P17, WORKING DRAFT SEQUENCE.
ACCESSION AC129445
VERSION AC129445.6 GI:30521811
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 242754)
Muzny,D,Marie., Metzker,M,Lee., Abramzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroil,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,

```

Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,I., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hughes,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokemele,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savary,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Vallas,R., Wang,V., Villalana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wlezyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,D., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhauser,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Unpublished Direct Submission

2 (bases 1 to 242754)

Worley,K.C.

Direct Submission

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

```

Assembly program: Atlas 3.0;
Consensus quality: 228124 bases at least Q40
Consensus quality: 230998 bases at least Q30
Consensus quality: 233152 bases at least Q20
Estimated insert size: 237219; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 242754: contig of 242754 bp in length.
Location/Qualifiers
1..242754
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-13P17"
524..1247
/note="clone boundary"
clone_end:Sp6
site:EcoRI
end_sequence:BH276372"
98074..99377
/note="wgs contig"
complement(234009..234650)
/note="clone boundary"
clone_end:T7
site:EcoRI
end_sequence:BH276370"
236507..237641
/note="wgs end_extension"
clone_end:T7"
241197..242754
/note="wgs end_extension"
clone_end:T7"

ORIGIN
Query Match 15.3%; Score 37.6; DB 2; Length 242754;
Best Local Similarity 48.6%; Pred. No. 5;
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 9 CTTTGCAGGATACCTTTATTTCTTTAAGATTCCTGTGTTTATACACAGATTTT 68
Db 211568 CCTGTGCTAAGAGTGTGTTGATTTTGTGTTTATTCAGTCAACTAAAAAGCTTCTGC 211627

QY 69 ACTTTACTCTACTGTGACCCAGTGAATTCCTTCTCAGTCACAGTGTCAACTCTA 128
Db 211628 CATTCCTGCCAATCTTAAACCTAGTGTGCTCTCTCAAAATAATGTTCTTAAAGCTG 211687

QY 129 CCCCCCACTGCAACAGAGTTTGTAGGGGCGATCAATCACCGAGAGTCAAGCCCT 188
Db 211688 TTCTGTAGTCCAGATAGACTTTATATGAGAACCAATATAGATTTTAAAGCCCTTGCCCT 211747

QY 189 CAACCACTAGGTGTGGGGGGGTAGGATCTG 220
Db 211748 GTACAGGTGGGCGAGTGTAAACAGTTAGCTG 211779

RESULT 13
AC105857/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-95G8, *** SEQUENCING IN PROGRESS ***,
ACCESSION AC105857
VERSION AC105857.4 GI:23101480

```

KEYWORDS SOURCE ORGANISM

HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
Rattus norvegicus (Norway rat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE AUTHORS

1 (bases 1 to 247480)
Muzny, D. Marie, Metzker, M. Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsebrook, S., Amin, A., Anguiano, D.,
Anylebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buha, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, J., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Devila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Frazer, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gebregorgis, B., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Guharathne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jollivet, A.,
Karpach, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorensu, H., Loulseg, H., Lozano, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwako, O., Okwuonu, G., Olarnpungoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smaj, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soia, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umani, K.,
Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
Weinstock, G. and Gibbs, R. A.

TITLE JOURNAL

Unpublished
2 (bases 1 to 247480)
Worley, K. C.

REFERENCE AUTHORS

Submitted (10-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

TITLE JOURNAL

Unpublished
3 (bases 1 to 247480)
Rat Genome Sequencing Consortium.

REFERENCE AUTHORS

Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Sep 18, 2002 this sequence version replaced gi:21736935.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas

(http://www.hgsc.bcm.tmc.edu/projects/rat/). As a result, the sequence may extend beyond the ends of the clone and there may be contigs that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNXX
Center clone name: CH230-95G8
----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 223628 bases at least Q40

Consensus quality: 227305 bases at least Q30

Consensus quality: 229716 bases at least Q20

Estimated insert size: 240475; sum-of-contigs estimation

Quality coverage: 5x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 243951: contig of 243951 bp in length

* 243952 244051: gap of unknown length

* 244052 247480: contig of 3429 bp in length.

FEATURES

source

1. 247480

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-95G8"

239..1070

/note="clone boundary"

clone_end:Sp5

site:EcoRI

end sequence:BH359863"

complement(237189..238065)

/note="clone boundary"

clone_end:T7

site:EcoRI

end sequence:BH359858"

239489..240989

/note="wgs_end_extension"

clone_end:T7

241930..243951

/note="wgs_end_extension"

clone_end:T7

244052..245191

/note="wgs_end_extension"

clone_end:T7

ORIGIN

Query Match 15.3%; Score 37.6; DB 2; Length 247480;
Best Local Similarity 48.6%; Pred No. 5;
Matches 103; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 9 CCTTGAAGATACCTTTTATTTCTTTAAAGTCTCTGTGTTTATACACAGATTTTA 68
|||||
Db 196343 CCTGTGCTAAGAGTGTTCGATTTGTTTATTTTATTCAGTCAACTAAAAGCTTCTGC 196284
|||||

Qy 69 AGTTTACTCTTACGTGACCCCAAGTGAATTCCTTCTCCAGTCACAGTGTCAACCTCTTA 128
|||||
Db 196283 CATTCCTGCCAATCTTCAAACTTAAGTGGTCTCTCTCTCAATAAATGGTTCTAAAGCTG 196224
|||||

Qy 129 CCCCCCACTGCACGAGAGTGTGTGAGGGGCATCATCACCCGAGAGTACAGAGCCCT 188
|||||
Db 196223 TTTCTTGTAGTCCAGATAGACTTTATTATGAGAACCATTAGATTTTAAAGCCCTTCGCCT 196164
|||||
Qy 189 CAACCACTGAGGTGTGTGGGGGGTAGGGATCTG 220
|||||
Db 196163 GTACAGGTGGGGCGAGGTGAACAGTTAGCTG 196132
|||||

RESULT 14

AC090127

LOCUS

DEFINITION

AC090127

AC090127

HTG.

VERSION

AC090127.11

GI:20330898

SOURCE

Mus musculus

ORGANISM

Mus musculus

(house mouse)

REFERENCE

1 (bases 1 to 212404)

2 (bases 1 to 212404)

3 (bases 1 to 212404)

4 (bases 1 to 212404)

5 (bases 1 to 212404)

6 (bases 1 to 212404)

7 (bases 1 to 212404)

8 (bases 1 to 212404)

9 (bases 1 to 212404)

10 (bases 1 to 212404)

11 (bases 1 to 212404)

12 (bases 1 to 212404)

13 (bases 1 to 212404)

14 (bases 1 to 212404)

15 (bases 1 to 212404)

16 (bases 1 to 212404)

17 (bases 1 to 212404)

18 (bases 1 to 212404)

19 (bases 1 to 212404)

20 (bases 1 to 212404)

21 (bases 1 to 212404)

22 (bases 1 to 212404)

23 (bases 1 to 212404)

24 (bases 1 to 212404)

25 (bases 1 to 212404)

26 (bases 1 to 212404)

27 (bases 1 to 212404)

28 (bases 1 to 212404)

29 (bases 1 to 212404)

30 (bases 1 to 212404)

31 (bases 1 to 212404)

32 (bases 1 to 212404)

33 (bases 1 to 212404)

34 (bases 1 to 212404)

35 (bases 1 to 212404)

36 (bases 1 to 212404)

37 (bases 1 to 212404)

38 (bases 1 to 212404)

39 (bases 1 to 212404)

40 (bases 1 to 212404)

41 (bases 1 to 212404)

42 (bases 1 to 212404)

43 (bases 1 to 212404)

44 (bases 1 to 212404)

45 (bases 1 to 212404)

46 (bases 1 to 212404)

47 (bases 1 to 212404)

48 (bases 1 to 212404)

49 (bases 1 to 212404)

50 (bases 1 to 212404)

51 (bases 1 to 212404)

52 (bases 1 to 212404)

53 (bases 1 to 212404)

54 (bases 1 to 212404)

55 (bases 1 to 212404)

56 (bases 1 to 212404)

57 (bases 1 to 212404)

58 (bases 1 to 212404)

59 (bases 1 to 212404)

60 (bases 1 to 212404)

61 (bases 1 to 212404)

62 (bases 1 to 212404)

63 (bases 1 to 212404)

64 (bases 1 to 212404)

65 (bases 1 to 212404)

66 (bases 1 to 212404)

67 (bases 1 to 212404)

68 (bases 1 to 212404)

69 (bases 1 to 212404)

70 (bases 1 to 212404)

71 (bases 1 to 212404)

72 (bases 1 to 212404)

73 (bases 1 to 212404)

74 (bases 1 to 212404)

75 (bases 1 to 212404)

76 (bases 1 to 212404)

77 (bases 1 to 212404)

78 (bases 1 to 212404)

79 (bases 1 to 212404)

80 (bases 1 to 212404)

81 (bases 1 to 212404)

82 (bases 1 to 212404)

83 (bases 1 to 212404)

84 (bases 1 to 212404)

85 (bases 1 to 212404)

86 (bases 1 to 212404)

87 (bases 1 to 212404)

88 (bases 1 to 212404)

89 (bases 1 to 212404)

90 (bases 1 to 212404)

91 (bases 1 to 212404)

92 (bases 1 to 212404)

93 (bases 1 to 212404)

94 (bases 1 to 212404)

95 (bases 1 to 212404)

96 (bases 1 to 212404)

97 (bases 1 to 212404)

98 (bases 1 to 212404)

99 (bases 1 to 212404)

100 (bases 1 to 212404)

101 (bases 1 to 212404)

102 (bases 1 to 212404)

103 (bases 1 to 212404)

104 (bases 1 to 212404)

105 (bases 1 to 212404)

106 (bases 1 to 212404)

107 (bases 1 to 212404)

108 (bases 1 to 212404)

109 (bases 1 to 212404)

110 (bases 1 to 212404)

111 (bases 1 to 212404)

112 (bases 1 to 212404)

113 (bases 1 to 212404)

114 (bases 1 to 212404)

115 (bases 1 to 212404)

116 (bases 1 to 212404)

117 (bases 1 to 212404)

118 (bases 1 to 212404)

119 (bases 1 to 212404)

120 (bases 1 to 212404)

121 (bases 1 to 212404)

122 (bases 1 to 212404)

123 (bases 1 to 212404)

124 (bases 1 to 212404)

125 (bases 1 to 212404)

126 (bases 1 to 212404)

127 (bases 1 to 212404)

128 (bases 1 to 212404)

129 (bases 1 to 212404)

130 (bases 1 to 212404)

131 (bases 1 to 212404)

132 (bases 1 to 212404)

133 (bases 1 to 212404)

134 (bases 1 to 212404)

135 (bases 1 to 212404)

136 (bases 1 to 212404)

137 (bases 1 to 212404)

138 (bases 1 to 212404)

139 (bases 1 to 212404)

140 (bases 1 to 212404)

141 (bases 1 to 212404)

142 (bases 1 to 212404)

143 (bases 1 to 212404)

144 (bases 1 to 212404)

145 (bases 1 to 212404)

146 (bases 1 to 212404)

147 (bases 1 to 212404)

148 (bases 1 to 212404)

149 (bases 1 to 212404)

150 (bases 1 to 212404)

151 (bases 1 to 212404)

152 (bases 1 to 212404)

153 (bases 1 to 212404)

154 (bases 1 to 212404)

155 (bases 1 to 212404)

156 (bases 1 to 212404)

157 (bases 1 to 212404)

158 (bases 1 to 212404)

159 (bases 1 to 212404)

160 (bases 1 to 212404)

161 (bases 1 to 212404)

162 (bases 1 to 212404)

163 (bases 1 to 212404)

164 (bases 1 to 212404)

165 (bases 1 to 212404)

166 (bases 1 to 212404)

167 (bases 1 to 212404)

168 (bases 1 to 212404)

169 (bases 1 to 212404)

170 (bases 1 to 212404)

171 (bases 1 to 212404)

172 (bases 1 to 2

JOURNAL

REFERENCE

AUTHORS

Submitted (26-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 212404)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lakoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., MacDonald, P., Major, J., Marguis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, K., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (27-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 26, 2002 this sequence version replaced gi:20219124.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L11594

Center clone name: 128_D_23

FEATURES

source

Location/Qualifiers

1..212404

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/chromosome="6"

/map="6"

/clone="RP23-128D23"

/clone_lib="RPCI-23 Female Mouse BAC"

7..50

/rpt_family="AT_rich"

/complement(528..841)

/rpt_family="L1"

855..1411

/rpt_family="L1"

1491..1527

/rpt_family="CA)n"

/complement(3065..3365)

/rpt_family="Lx8"

/complement(4852..5300)

/rpt_family="Lx8"

5395..5436

/rpt_family="Lx8"

5447..5491

/rpt_family="TTTA)n"

/rpt_family="CT-rich"

/complement(5509..5795)

/rpt_family="Lx8"

/complement(5799..5851)

/rpt_family="B1_MM"

7807..7982

/rpt_family="TTTC)n"

/complement(7995..8272)

repeat_region /rpt_family="MTD" 8281..8302
repeat_region /rpt_family="AT_rich" 8312..8353
repeat_region /rpt_family="CA)n" 8355..8536
repeat_region /rpt_family="GA-rich" complement(8537..10801)
unsure /rpt_family="Lx8" complement(9385..9389)
unsure /note="<30 qual SNGL region" complement(9492..9497)
repeat_region 11578..12353
repeat_region /rpt_family="L1" 12354..12466
repeat_region /rpt_family="CA)n" 12467..13522
unsure /rpt_family="L1" 13160..13165
unsure /note="<30 qual SNGL region" 13204..13208
repeat_region /note="<30 qual SNGL region" 14255..14463
repeat_region /rpt_family="L1" complement(15279..15586)
repeat_region /rpt_family="Lx8" 15747..16085
repeat_region /rpt_family="ORRIA3" 16086..16201
repeat_region /rpt_family="ORRIA-int" complement(16258..16997)
repeat_region /rpt_family="Lx" 17042..17064
repeat_region /rpt_family="CA)n" 17406..17440
repeat_region /rpt_family="AT_rich" complement(17453..18422)
repeat_region /rpt_family="Lx2" 18423..18816
repeat_region /rpt_family="L1" 21111..21130
repeat_region /rpt_family="CA)n" 21131..21157
repeat_region /rpt_family="CA)n" complement(21862..22008)
repeat_region /rpt_family="B1_MM" 22363..22399
repeat_region /rpt_family="AT_rich" 22719..23842
repeat_region /rpt_family="L1_MM" 23878..24271
repeat_region /rpt_family="MTC" 24703..24799
repeat_region /rpt_family="TCTA)n" complement(25422..25540)
repeat_region /rpt_family="Lx8" complement(25692..25756)
repeat_region /rpt_family="B1F" 25762..25782
repeat_region /rpt_family="AT_rich" complement(25783..27197)
repeat_region /rpt_family="Lx" 27265..27368
repeat_region /rpt_family="Lx5"

Query Match 15.2%; Score 37.4; DB 10; Length 212404;

Best Local Similarity 60.2%; Pred. No. 5.7;

Matches 62; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 5 TTAACCTTTCGAAGATACCTTTTATTTCTTTAAAGATTCCTGTTTATATACACAGAT 64

DB 17395 TTTAAGTTGCTATATATATCTTTTAAAGTATTATTTTGTATTACTTAGTT 17454

QY 65 TTTAAGTTTACTCTACTGCTGACCAAGTGAATTCCTTC 107
|||||
Db 17455 TTTTATTACATCAACACGCTTAAACCACTCCCAAGTCCATTCTC 17497
|||||

RESULT 15

AC138731/c

LOCUS

DEFINITION

AC138731

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

REFERENCE

AUTHORS

JOURNAL

AC138731 160699 bp DNA linear HTG 28-JAN-2003
Pongo pygmaeus clone CH253-404N12, WORKING DRAFT SEQUENCE, 8
ordered pieces.

AC138731.2 GI:279233655
HTG: HTGS PHASE2; HTGS DRAFT.

Pongo pygmaeus (orangutan)

Pongo pygmaeus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.

1 (bases 1 to 160699)

Akhter N., Antonellis, A., Ayele, K., Beckstrom-Sternberg, S.M.,

Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,

Carlaga, K., Coleman, B., Engle, J., Granite, S., Guan, X., Gupta, J.,

Haghghi, P., Han, J., Hansen, N., Ho, S.-L., Idol, J.R., Karlins, E.,

Laric, P., Lee-Lin, S.-Q., Legaspi, R., Maduro, Q.L., Maduro, V.B.,

Marques, E.H., Masiello, C., Maskeri, B., McDowell, J.,

Paguirigan, C., Pearson, R., Portnoy, M.E., Prasad, A.,

Reddix-Dugue, N., Schandler, K., Schueler, M.G., Sison, C.,

Statropop, S., Thomas, J.W., Thomas, P.J., Touchman, J.W., Vogt, J.L.,

Wetherby, K.D., Wiggins, L., Young, A., and Green, E.D.

NISC Comparative Sequencing Initiative

Unpublished

2 (bases 1 to 160699)

Green, E.D.

Direct Submission

Submitted (16-JAN-2003) NIH Intramural Sequencing Center, 8717

Groveomont Circle, Gaithersburg, MD 20877, USA

3 (bases 1 to 160699)

Green, E.D.

Direct Submission

Submitted (28-JAN-2003) NIH Intramural Sequencing Center, 8717

Groveomont Circle, Gaithersburg, MD 20877, USA

On Jan 28, 2003 this sequence version replaced gi:27764645.

----- Genome Center

Center: NIH Intramural Sequencing Center

Center code: NISC

Web site: <http://www.nisc.nih.gov>

Contact: nisc_zoombgri.nih.gov

----- Project Information

Center project name: ela

Center clone name: 404N12

The sequence data in this record represents an 'enhanced'

version of a Phase 2 submission. Specifically, the indicated

order and orientation of each sequence contig has been

established using one or more of the following: read-pair

data from individual subclones, overlaps with neighboring

clones, alignment with available reference sequence (e.g.,

human), and/or confirmation by PCR testing. In addition,

the sequence assembly is based on at least 8X average

coverage in Q20 bases and has been reviewed to rule out

gross misassemblies, the low-quality ends of sequence

contigs have been trimmed away, and each base is associated

with a Phrap-derived quality score.

----- Summary Statistics

Sequencing vector: plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 159180 bases at least Q40

Consensus quality: 159652 bases at least Q30

Consensus quality: 159822 bases at least Q20

Insert size: 147000; agarose-fp

Insert size: 159999; sum-of-contigs

Quality coverage: 11.17x in Q20 bases; agarose-fp

Quality coverage: 10.26x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 8 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 27629: contig of 27629 bp in length

* 27630 27729: gap of unknown length

* 27730 59021: contig of 31292 bp in length

* 59022 59121: gap of unknown length

* 59122 63118: contig of 4797 bp in length

* 63119 64018: gap of unknown length

* 64019 91239: contig of 27221 bp in length

* 91240 91339: gap of unknown length

* 91340 125098: contig of 33759 bp in length

* 125099 125198: gap of unknown length

* 125199 132510: contig of 7312 bp in length

* 132511 132610: gap of unknown length

* 132611 144517: contig of 11907 bp in length

* 144518 144617: gap of unknown length

* 144618 160699: contig of 16082 bp in length.

Location/Qualifiers

1..160699

/organism="Pongo pygmaeus"

/mol_type="genomic DNA"

/db_xref="taxon:9600"

/clone="CH253-404N12"

/clone_lib="CH253"

1..27629

/note="assembly_fragment"

clone end:SP6

vector_side:left

27730..59021

/note="assembly_fragment"

59122..63118

/note="assembly_fragment"

64019..91239

/note="assembly_fragment"

91340..125098

/note="assembly_fragment"

125199..132510

/note="assembly_fragment"

132611..144517

/note="assembly_fragment"

144618..160699

/note="assembly_fragment"

clone end:T7

vector_side:right

ORIGIN

Query Match 15.1%; Score 37.2; DB 2; Length 160699;

Best Local Similarity 56.6%; Pred. No. 6.5;

Matches 69; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 2 TCATTAACTTTCAGGATACCTTTTATTTCTTTAAGATTCCTGTTTATACACA 61

Db 89101 TAACTAAGCTTCTCAAGGCTATCTTCTATCTTCACTATATCTTTTCTACTTAATCCCT 89042

QY 62 GATTTTAACTTCTCTCTCTGCTGACCAAGTGAATTCCTTCTCCAGTCACAGTGTCA 121

Db 89041 GATATTCAAAATATATTAAAGAGGTGGAGGAGGAATGCTTCCCATGTACATTATA 88982

QY 122 AC 123

Db 88981 AC 88980

Search completed: June 10, 2004, 18:22:48

Job time : 1719 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 15:30:58 ; Search time 344 Seconds

(without alignments)
3037.956 Million cell updates/sec

Title: US-09-674-593-10

Perfect score: 246
Sequence: 1 gtcattacaccttgcaagga.....ttcatatcacccccacacta 246

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq 29Jan04: *
1: geneseqn1980s: *
2: geneseqn1990s: *
3: geneseqn2000s: *
4: geneseqn2001as: *
5: geneseqn2001bs: *
6: geneseqn2002s: *
7: geneseqn2003as: *
8: geneseqn2003bs: *
9: geneseqn2003cs: *
10: geneseqn2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	246	100.0	246	3	Aaz36649 EST AA863
2	244.4	99.3	1382	3	Aaz36643 Human tum
3	38.4	15.6	242	4	AAL10336 Human bre
C 4	36.2	14.7	3588	7	ACA28683 Prokaryot
5	35.6	14.5	507	3	Aaz80742 Human col
6	35.6	14.5	24387	4	AAL07156 Human rep
C 7	35.4	14.4	2803	6	ABK35341 Human cdn
C 8	35.4	14.4	3644	4	AAC94450 Human cdn
C 9	35.4	14.4	3710	4	AAH17830 Human cdn
C 10	35.4	14.4	4480	6	ABK28656 Human cdn
C 11	34.8	14.1	2264	9	ADK53917 Human pro
C 12	34.4	14.0	2239	4	ABL03611 Drosophil
C 13	34.4	14.0	4405	4	ABL03610 Drosophil
C 14	34.4	14.0	12860	4	ABL03530 Drosophil
C 15	34.2	13.9	388	4	AAK55128 Human imm
C 16	34	13.8	17213	6	ABL33482 Human imm
C 17	34	13.8	110000	6	ABQ74964 ⁶ Continuation (7 of
C 18	33.2	13.5	3759	7	ACA28109 Prokaryot
C 19	32.8	13.3	16560	4	ABL17922 Drosophil
C 20	32.8	13.3	19165	4	ABL17898 Drosophil
C 21	32.4	13.2	580	6	ABQ59302 Human col
C 22	32.2	13.1	15674	6	ABL32362 Human imm
23	32.2	13.1	15674	6	ABL34476 Human met

24	32.2	13.1	15674	6	ABL70513	Chemical1
C 25	32.2	13.1	98690	6	ABK12169	Human DNA
C 26	32	13.0	480	7	ACA27596	Prokaryot
C 27	32	13.0	2175	3	AAC79978	Human sec
C 28	31.8	12.9	422	8	ACH15465	Human adu
C 29	31.8	12.9	1885	3	AAA50339	Human myz
C 30	31.8	12.9	4346	9	ADE25645	Human cdn
C 31	31.6	12.8	472	8	ACH39321	Human foe
C 32	31.6	12.8	700	4	AAH92432	Human inf
C 33	31.6	12.8	2337	8	ADA32494	DNA encod
C 34	31.6	12.8	13158	2	AAT75288	Nucleotid
C 35	31.4	12.8	1017	2	AAK99542	Nucleic a
C 36	31.4	12.8	1215	6	ABL90038	Human pol
C 37	31.4	12.8	8467	6	ABL32108	Human imm
C 38	31.4	12.8	8758	6	ABL33119	Human imm
C 39	31.2	12.7	400	7	ABX43460	Bovine ES
C 40	31.2	12.7	437	4	AAK56796	Human imm
C 41	31.2	12.7	458	7	ABZ18171	Group III
C 42	31.2	12.7	468	4	AAI10244	Probe #17
C 43	31.2	12.7	468	4	AEA51880	Human foe
C 44	31.2	12.7	468	4	AAI31492	Probe #17
C 45	31.2	12.7	468	4	AEA21701	Probe #16

ALIGNMENTS

RESULT 1

AAZ36649
ID AAZ36649 standard; cDNA; 246 BP.

XX
AC AAZ36649;

XX
DT 22-FEB-2000 (first entry)

XX
DE EST AA863443 which overlaps the RUR-1 antisense cDNA sequence.

XX
KW Antisense; human; tumour rejection antigen; RUR-1; tumour; cancer;

XX
KW renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma; leukaemia;

XX
KW EST; expressed sequence tag; ss.

XX
OS Homo sapiens.

XX
PN WO9558546-A1.

XX
PD 18-NOV-1999.

XX
PF 13-MAY-1999; 99WO-US010424.

XX
PR 13-MAY-1998; 98US-0085318P.

XX
(LUDW-) LUDWIG INST CANCER RES.

XX
Van Den Eynde B, Boon-Falleur T;

XX
WPI; 2000-053076/04.

XX
PT New isolated tumor rejection antigen RUR-1 nucleic acids, used for, e.g.

XX
PS treatment of cancers.

XX
Claim 8; Page 71; 75pp; English.

XX
The present sequence represents an expressed sequence tag (EST) which overlaps the antisense cDNA sequence of human tumour rejection antigen RUR-1. The RUR-1 antisense sequence is the antisense strand of a ubiquitously expressed gene. The antisense strand codes for a polypeptide which is preferentially expressed in tumour samples and tumour-derived cell lines. The polypeptide is unrelated to any TRAP protein. The RUR-1 sequence was isolated from a renal cell carcinoma line LB9211-RCC. The RUR-1 nucleic acids and polypeptides can be used for diagnosis, prognosis or treatment of a disorder characterized by the expression of a RUR-1 antisense cDNA molecule or an expression product, such as cancers, e.g. renal cell carcinoma, colorectal carcinoma, melanoma, sarcoma or

CC leukaemia. note: although the present sequence is mentioned in claim 8,
CC it is not specifically claimed. The fragments of AAZ36643-44 which DO NOT
CC contain the present sequence are claimed
XX
SQ Sequence 246 BP; 63 A; 66 C; 44 G; 73 T; 0 U; 0 Other;
Query Match 100.0%; Score 246; DB 3; Length 246;
Best Local Similarity 100.0%; Pred. No. 7.3e-65;
Matches 246; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCATTAACTTTGCAAGATACCTTTTATTTCTTTAAAGATTCTGTGTTTATACAC 60
Db 1 GTCATTAACTTTGCAAGATACCTTTTATTTCTTTAAAGATTCTGTGTTTATACAC 60
QY 61 AGATTTTAAAGTTTACTCTTACTGCTGACCCCAAGTGAATTCCTTCCAGTCACAGTGTC 120
Db 61 AGATTTTAAAGTTTACTCTTACTGCTGACCCCAAGTGAATTCCTTCCAGTCACAGTGTC 120
QY 121 AACCTCTACCCCCCACTGCAACGAGAGTTTGTAGGGGCATCAATCACACCGAGAAGTCA 180
Db 121 AACCTCTACCCCCCACTGCAACGAGAGTTTGTAGGGGCATCAATCACACCGAGAAGTCA 180
QY 181 CAGCCCTCAACACATGAGGTGTGGGGGGTGGGATCTGCAATTCCTCATATCAACCCC 240
Db 181 CAGCCCTCAACACATGAGGTGTGGGGGGTGGGATCTGCAATTCCTCATATCAACCCC 240
QY 241 ACACCTA 246
Db 241 ACACCTA 246
RESULT 2
AAZ36643
ID AAZ36643 standard; cDNA; 1382 BP.
XX
AC AAZ36643;
XX
DT 22-FEB-2000 (first entry)
XX
DE Human tumor rejection antigen RUR-1 antisense cDNA sequence.
XX
KW Antisense; human; tumor rejection antigen; RUR-1; tumour; cancer;
KW renal cell carcinoma; colorectal carcinoma; melanoma; sarcoma; leukaemia;
KW ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT primer_bind 523..547
FT /*tag= b
FT /*note= "binding site for primer VDE119 (see AAZ36647)"
FT CDS 738..992
FT /*tag= a
FT primer_bind complement(1280..1305)
FT /*tag= c
FT /*note= "binding site for primer VDE120 (see AAZ36648)"
XX
PN WO9558546-A1.
XX
PD 18-NOV-1999.
XX
PF 13-MAY-1999; 99WO-US010424.
XX
PR 13-MAY-1998; 98US-0085318P.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Van Den Eynde B, Boon-Falleur T;
XX
XX WPI; 2000-053076/04.
DR P-PSDB; AAY53809.
XX
XX New isolated tumor rejection antigen RUR-1 nucleic acids, used for, e.g.

PT treatment of cancers.
XX
PS Claim 4; Fig 5; 75pp; English.
XX
CC The present sequence represents the antisense cDNA sequence of human
CC tumor rejection antigen RUR-1. The present sequence is the antisense
CC strand of a ubiquitously expressed gene. The antisense strand codes for a
CC polypeptide which is preferentially expressed in tumour samples and
CC tumour-derived cells lines. The polypeptide is unrelated to any TRAP
CC protein. The sequence was isolated from a renal cell carcinoma line
CC IB9211-RCC. The RUR-1 nucleic acids and polypeptides can be used for
CC diagnosis, prognosis or treatment of a disorder characterized by the
CC expression of a RUR-1 antisense cDNA molecule or an expression product,
CC such as cancers, e.g. renal cell carcinoma, colorectal carcinoma,
CC melanoma, sarcoma or leukaemia
XX
SQ Sequence 1382 BP; 355 A; 373 C; 344 G; 310 T; 0 U; 0 Other;
Query Match 99.3%; Score 244.4; DB 3; Length 1382;
Best Local Similarity 99.6%; Pred. No. 4.2e-64;
Matches 245; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GTCATTAACTTTGCAAGATACCTTTTATTTCTTTAAAGATTCTGTGTTTATACAC 60
Db 268 GTCATTAACTTTGCAAGATACCTTTTATTTCTTTAAAGATTCTGTGTTTATACAC 327
QY 61 AGATTTTAAAGTTTACTCTTACTGCTGACCCCAAGTGAATTCCTTCCAGTCACAGTGTC 120
Db 328 AGATTTTAAAGTTTACTCTTACTGCTGACCCCAAGTGAATTCCTTCCAGTCACAGTGTC 387
QY 121 AACCTCTACCCCCCACTGCAACGAGAGTTTGTAGGGGCATCAATCACACCGAGAAGTCA 180
Db 388 AACCTCTACCCCCCACTGCAACGAGAGTTTGTAGGGGCATCAATCACACCGAGAAGTCA 447
QY 181 CAGCCCTCAACACATGAGGTGTGGGGGGTGGGATCTGCAATTCCTCATATCAACCCC 240
Db 448 CAGCCCTCAACACATGAGGTGTGGGGGGTGGGATCTGCAATTCCTCATATCAACCCC 507
QY 241 ACACCTA 246
Db 508 ACACCTA 513
RESULT 3
AAL10336/c
ID AAL10336 standard; cDNA; 242 BP.
XX
AC AAL10336;
XX
DT 07-DEC-2001 (first entry)
XX
DE Human breast cancer expressed polynucleotide 2793.
XX
KW Human; breast cancer; cell marker; cytostatic; ss.
XX
OS Homo sapiens.
XX
PN WO200151628-A2.
XX
PD 19-JUL-2001.
XX
PF 10-JAN-2001; 2001WO-US000798.
XX
PR 14-JAN-2000; 2000US-0176077P.
PR 14-MAR-2000; 2000US-0189167P.
PR 24-MAR-2000; 2000US-0192099P.
PR 29-MAR-2000; 2000US-0193480P.
PR 15-MAY-2000; 2000US-0205230P.
PR 09-JUN-2000; 2000US-0211315P.
PR 25-JUL-2000; 2000US-0220534P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX

PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX WPI; 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer.
XX
XX Claim 1; Page 524; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AA107544-AA126789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterizing treating and
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity
XX
SQ Sequence 242 BP; 89 A; 56 C; 62 G; 35 T; 0 U; 0 Other;
Query Match 15.6%; Score 38.4; DB 4; Length 242;
Best Local Similarity 56.2%; Pred. No. 0.12; Mismatches 0; Gaps 0;
Matches 72; Conservative 0; Indels 56; Indels 0; Gaps 0;
QY 24 CTTTATTATTTCTTTAAGATTCCTGTTGTTTATACACAGATTTTAAGTTTACTCTACTG 83
DB 231 CTTTCGTTGGTTTCTTTTCTTTTCTTTTCTATATATTTTGTGATTCUATTT 172
QY 84 CTGACCCCAAGTGAATTCCTTCTCCAGTCCACAGTGTCAACCTTACCCCCCACTGCAAC 143
DB 171 TATTTTAAATCTCTCTCTCTCCAGACAAATGGCACTGCTTATCTCCGAAATGGTGT 112
QY 144 GAGAGTTT 151
DB 111 GATCGTCT 104
RESULT 4
ACA28683/c
ID ACA28683 standard; DNA; 3588 BP.
XX
AC ACA28683;
XX
DT 19-JUN-2003 (first entry)
DE Prokaryotic essential gene #10340.
XX
XX Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX
OS Clostridium botulinum.
XX
FN WO20027183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
XX (BLIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR P-PSDB; ABU24813.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 14; SEQ ID NO 16553; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 3588 BP; 1722 A; 290 C; 624 G; 952 T; 0 U; 0 Other;
Query Match 14.7%; Score 36.2; DB 7; Length 3588;
Best Local Similarity 56.2%; Pred. No. 1.4;
Matches 68; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
QY 2 TCATTAAACCTTTGCAAGGATACCTTTTATTTTCTTTTAAAGATTCCTGTTTATACACA 61
DB 3095 TCATAACATTCAAAAGTTCCTTTTGACTTTTACTAAATCTTCTTTGATTATTCATA 3036
QY 62 GATTTTAAAGTTTACTCTCTGACCCCAAGTGAATTCCTTCTCCAGTCACAGTGCA 121
DB 3035 AATGTTATTTTTCATTAAATTCCTTATATTCCTTATTCCTTATTCCTTATTCGTGCT 2976
QY 122 A 122
DB 2975 A 2975
RESULT 5
AAZ80742
ID AAZ80742 standard; cDNA; 507 BP.
XX
AC AAZ80742;
XX
DT 07-APR-2000 (first entry)
XX
DE Human colon cancer cell line SW480 cDNA clone SEQ ID NO:826.
XX
XX Human; gene expression product; diagnosis; tumour; colon cancer;
KW colorectal adenocarcinoma; cell line SW480; cell proliferation;
KW cytostatic; sarcoma; breast cancer; neoplasia; dysplasia; hyperplasia;
XX ds.
OS Homo sapiens.
XX

```
PN WO9964576-A2.
XX 16-DEC-1999.
XX 09-JUN-1999; 99WO-IB001062.
XX 10-JUN-1998; 98US-0088801P.
XX (FARB ) BAYER CORP.
XX Endege WO, Steinmann KE, Astle JH, Burgess CC, Bushnell SE;
PI Carroll E, Catino TJ, Derti A, Ford DM, Lewis ME, Monahan JB;
PI Schlegel R;
XX WPI; 2000-087220/07.
XX Novel nucleic acids, used to develop products for the diagnosis and
PT treatment of disorders involving unwanted cell proliferation,
PT particularly cancers, especially colon cancer.
XX Claim 15; Page 459; 469pp; English.
XX AA279917 to AAZ80766 represent double stranded cDNA clones isolated from
CC the human colorectal adenocarcinoma (colon cancer) cell line SW480. The
CC cDNA clones can be used to generate antisense oligonucleotides which can
CC be used for antisense therapy. Methods and products from the present
CC invention can be used for identifying and/or classifying cancerous cells
CC present in a human tumour, particularly in solid tumours, e.g. carcinomas
CC and sarcomas, e.g. breast or colon cancers. The cDNA clones can be used
CC for developing agents for the diagnosis and treatment of disorders
CC involving unwanted cell proliferation, such as neoplasia, dysplasia or
CC hyperplasia
XX SQ Sequence 507 BP; 132 A; 148 C; 71 G; 152 T; 0 U; 4 Other;
SQ Query Match 14.5%; Score 35.6; DB 3; Length 507;
Best Local Similarity 62.2%; Pred. No. 1.1;
Matches 56; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
QY 62 GATTTTAAGTTTACCTCTACTCTGACCCCAAGTGAATTCCTTCACAGTCACAGTCA 121
DB 203 GATTTTAATATATACACCTCAGGACCAAAGAAAAGTTAAGCAGAGGGTTCCA 262
QY 122 ACCTCTACCCCACTGCAACGAGAGTTT 151
DB 263 AGTGCTCTCTCCCACTTCACACAGATGT 292
RESULT 6
AAL07156
ID AAL07156 standard; DNA; 24387 BP.
XX AC AAL07156;
XX 21-NOV-2001 (first entry)
XX Human reproductive system related antigen DNA SEQ ID NO: 9844.
XX Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX Homo sapiens.
XX WO20015320-A2.
XX 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US001339.
XX 31-JAN-2000; 2000US-0179065P.
XX 04-FEB-2000; 2000US-0180628P.
XX 24-FEB-2000; 2000US-0184664P.
XX 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205151P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 14-AUG-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0234223P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234274P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
```

13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX Isolated nucleic acid molecule encoding a reproductive system antigen is
XX used in preventing, treating or ameliorating a medical condition.
PT
XX Disclosure; SEQ ID NO 9844; 1297pp + Sequence Listing; English.
PS
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
CC

CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention
XX
SQ Sequence 24387 BP; 5866 A; 6426 C; 6438 G; 5657 T; 0 U; 0 Other;
Query Match 14.5%; Score 35.6; DB 4; Length 24387;
Best Local Similarity 62.2%; Pred. No. 4.4;
Matches 56; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
QY 62 GATTTTAAGTTTACTCTACTCTGCTACCCCAAGTGAATTCCTTCTCCAGTCACAGTGCA 121
Db 19467 GATTTTAAATATATACACCTCAGGACCAAGAAAAAGTTAAGCAAGCGGGTTCCA 19526
QY 122 ACCTCTACCCCAACTGCAACGAGAGTTT 151
Db 19527 AGTGCTCTCTCCCACTTCAACAAGATGT 19556
RESULT 7
ABK35341/c
ID ABK35341 standard; cDNA; 2803 BP.
XX
AC ABK35341;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human cDNA encoding secreted protein #479.
XX
KW Human; secreted protein; gene; ss; nutritional supplement; haemophilia;
KW viral infection; bacterial infection; fungal infection; diabetes; asthma;
KW autoimmune disorder; rheumatoid arthritis; multiple sclerosis; tumour;
KW autoimmune thyroiditis; allergic reaction; neurodegenerative disease;
KW Alzheimer's disease; Parkinson's disease; liver fibrosis; cancer; ulcer;
KW coagulation disorder; inflammatory disorder; Crohn's disease; incision;
KW tissue regeneration; wound healing; burn; haematopoiesis;
KW myeloid cell deficiency; lymphoid cell deficiency.
XX
OS Homo sapiens.
XX
XX WO200177288-A2.
XX
XX 18-OCT-2001.
XX
XX 29-MAR-2001; 2001WO-US010224.
XX
XX 06-APR-2000; 2000US-0195582P.
XX
XX (GEMY) GENETICS INST INC.
XX
XX Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
XX Gulukota K, Graham JR;
XX
XX WPI; 2002-179321/23.
XX
XX Five hundred and ninety two polynucleotides derived from a variety of
XX human tissue sources which encode secreted proteins, useful for treating
XX immune deficiencies and disorders such as autoimmune disorders.
XX
XX Claim 1; Page 317-318; 372pp; English.
XX
XX The invention relates to 592 polynucleotides which have been derived from
XX a variety of human tissue sources and which encode novel secreted
XX proteins. The polynucleotides can be used as probes for the
XX identification and isolation of full length cDNA and genomic DNA. The
XX polynucleotides and proteins can also be used as nutritional supplements.
XX The proteins are useful in the treatment of various immune deficiencies
XX and disorders such as viral infections, bacterial infections, fungal
XX infections, autoimmune disorders (e.g. rheumatoid arthritis, multiple
XX sclerosis, autoimmune thyroiditis and diabetes) and allergic reactions
XX and conditions (e.g. asthma). They are also useful for treating
XX neurodegenerative diseases (e.g. Alzheimer's disease, Parkinson's
XX disease), liver fibrosis, coagulation disorders (e.g. haemophilia),
XX

CC inflammatory disorders (e.g. Crohn's disease) and tumours. They are also
CC useful for tissue regeneration, for wound healing and in the treatment of
CC burns, incisions and ulcers. The proteins are also useful for regulating
CC haematopoiesis and for treating myeloid or lymphoid cell deficiencies.
CC Sequences ABK34863-ABK35454 represent polynucleotides of the invention
XX
SQ Sequence 2803 BP; 688 A; 760 C; 712 G; 643 T; 0 U; 0 Other;

Query Match 14.4%; Score 35.4; DB 6; Length 2803;
Best Local Similarity 63.5%; Pred. No. 2.3;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 2 TCATTAACTTTGCAAGGATACCTTTTATTTCTTTAAGATTCCTGTTGTTATACACA 61
Db 2102 TCATTTTCCCATCCAGGATAGATATATATTTCTTTCATATTTATATATATATATA 2043
QY 62 GATTTTAAGTTTACTCCTACTGCTG 86
Db 2042 TATATTATATGTACACACACCTG 2018

RESULT 8
AAC84450/C
ID AAC84450 standard; cDNA; 3644 BP.
XX AAC84450;
XX
DT 02-APR-2001 (first entry)
DE
DE Nucleotide sequence of mddt cDNA clone ID No: 244366.6.
XX Molecule for disease detection and treatment; mddt; human; cirrhosis;
KW cell proliferative disorder; arteriosclerosis; bursitis; hepatitis;
KW connective tissue disease; myelofibrosis; cancer; leukemia; autoimmune;
KW acquired immunodeficiency syndrome; AIDS; Addison's disease; allergy;
KW anemia; bronchitis; gout; Hashimoto's thyroiditis; multiple sclerosis;
KW cytostatic; immunomodulator; anti-inflammatory; gene therapy; ss.
XX
OS Homo sapiens.
XX
XX WO200075298-A2.
XX 14-DEC-2000.
XX 01-JUN-2000; 2000WO-US015344.
XX 03-JUN-1999; 99US-0137412P.
PR 05-AUG-1999; 99US-0147500P.
PR 05-AUG-1999; 99US-0147501P.
PR 05-AUG-1999; 99US-0147542P.
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Hodgson DM, Lincoln SE, Russo FD, Spiro PA, Banville SC;
PI Bratcher SR, Dufour GE, Cohen HJ, Rosen BH, Hillman JL;
PI Jones AL, Yu JY, Greenawalt LB, Panzer SR, Roseberry AM, Wright RJ;
PI Daniels SE;
XX
XX WPI; 2001-071068/08.
XX
XX New polypeptide and polynucleotide molecules for disease detection and
PT treatment are useful in diagnosis and gene therapy of proliferative
PT disorders (e.g. breast cancer), autoimmune and inflammatory disorders
PT (e.g. AIDS or allergy).
XX
XX Claim 1; Page 93-94; 99pp; English.
XX
XX Sequences AAC84445-C84458 are new isolated polynucleotide molecules for
CC disease detection and treatment (mddt). The mddt polynucleotides are
CC useful for diagnosing or treating disorders associated with disease
CC detection and treatment molecules. These diseases include cell
CC proliferative disorders (e.g. arteriosclerosis, bursitis, cirrhosis,
CC hepatitis, mixed connective tissue disease, myelofibrosis, or cancers

CC such as leukemia, or breast or brain cancers), autoimmune or inflammatory
CC disorders (e.g. acquired immunodeficiency syndrome (AIDS), Addison's
CC disease, allergies, anemia, autoimmune haemolytic anemia, bronchitis,
CC gout, Hashimoto's thyroiditis, multiple sclerosis or cancers). The mddt
CC are especially useful for somatic or germline gene therapy. The mddt may
CC also be used to detect the presence of, or to quantify the amount of, an
CC mddt-related polynucleotide in a sample. The mddt are also useful for
CC isolating full length cDNA sequences utilizing hybridization and/or
CC amplification procedures, for generating hybridization probes useful in
CC chromosomal mapping of naturally occurring genomic sequences, as
CC molecular weight markers, or for monitoring the progress of disorders
CC associated with abnormal levels of mddt expression or evaluating the
CC efficacy of a particular treatment. The recombinant nucleic acids are
CC useful as part of a viral vector (e.g. based on a vaccinia virus) for
CC vaccinating a mammal and inducing a protective immunological response in
CC the mammal. The MDDT polypeptides are useful for screening molecules that
CC bind to or are bound by the polypeptides, i.e. for screening agonists,
CC antagonists or modulators of MDDT
XX
SQ Sequence 3644 BP; 856 A; 979 C; 897 G; 912 T; 0 U; 0 Other;

Query Match 14.4%; Score 35.4; DB 4; Length 3644;
Best Local Similarity 63.5%; Pred. No. 2.5;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 2 TCATTAACTTTGCAAGGATACCTTTTATTTCTTTAAGATTCCTGTTGTTATACACA 61
Db 2938 TCATTTTCCCATCCAGGATAGATATATATTTCTTTCATATTTATATATATATA 2879
QY 62 GATTTTAAGTTTACTCCTACTGCTG 86
Db 2878 TATATTATATGTACACACACCTG 2854

RESULT 9
AAH17830/C
ID AAH17830 standard; cDNA; 3710 BP.
XX AAH17830;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human cDNA sequence SEQ ID NO:17508.
DE
DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KW
XX Homo sapiens.
OS
XX EP1074617-A2.
PN
XX 07-FEB-2001.
PD
XX 28-JUL-2000; 2000EP-00116126.
PP
XX 29-JUL-1999; 99JP-00248036.
PR
XX 27-AUG-1999; 99JP-00300253.
PR
XX 11-JAN-2000; 2000JP-00118776.
PR
XX 02-MAY-2000; 2000JP-00183767.
PR
XX 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
DR
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
XX Claim 8; SEQ ID NO 17508; 2537pp + Sequence Listing; English.
PS

XX CC The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95993 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX SQ Sequence 3710 BP; 866 A; 999 C; 905 G; 940 T; 0 U; 0 Other;

Query Match 14.4%; Score 35.4; DB 4; Length 3710;
Best Local Similarity 63.5%; Pred. No. 2.6;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 2 TCATTAACTTTCAGAGTACCTTTTATTTTCTTTAAGATTCCTGTTTATACACA 61
DB 3009 TCATTTTCCATCCAGGATAGATATATATTTCTTTGATTTTAAATATATATATA 2950

QY 62 GATTTTAAAGTTTACTCCTACTGCTG 86
DB 2949 TATATTATATGTACACACACCTG 2925

RESULT 10
ID ABK28656/c
XX ABK28656 standard; cDNA; 4480 BP.
XX AC ABK28656;
XX DT 09-APR-2002 (first entry)
XX DE Human cDNA encoding secreted protein SECP23.
XX KW Human; ss; gene; SECP; antiinflammatory; cytostatic; cardiac; immunosuppressive; antiviral; anti-HIV; antiarthritic; antirheumatic; muscular active general; anticonvulsant; nootropic; neuroprotective; allergic; hypotensive; cardiovascular disorder; atherosclerosis; hypertension; myocardial infarction; autoimmune disorder; inflammatory disorder; AIDS; acquired immunodeficiency syndrome; allergy; rheumatoid arthritis; cell proliferative disorder; cancer; developmental disorder; Duchenne muscular dystrophy; neurological disorder; epilepsy; Alzheimer's disease.
XX OS Homo sapiens.
XX PN W0200198353-A2.
XX PD 27-DEC-2001.
XX PF 20-JUN-2001; 2001WO-US019862.
XX PR 20-JUN-2000; 2000US-0212890P.
XX PR 23-JUN-2000; 2000US-0213466P.
XX PR 27-JUN-2000; 2000US-0214601P.
XX PR 31-JUL-2000; 2000US-0223727P.
XX PR 08-SEP-2000; 2000US-0231435P.
XX PR 15-SEP-2000; 2000US-0232899P.

(INCY-) INCYTE GENOMICS INC.

XX PA Hillman JL, Tang YT, Yue H, Elliott VS, Tribouley CM, Lee EA;
XX PI Ramkumar J, Lal P, Xu Y, Warren BA, Hafalia AJA, Baughn MR;
XX PI Azimzai Y, Batra S, Burford N, Yao MG, Nguyen DB, Lu DAM, Walla NK;
XX PI Gandhi AR, Au-Young J, Patterson C;
XX DR WPI; 2002-090431/12.
XX DR P-PSDB; AAU81997.

XX PT Forty four human secreted proteins (referred to as SECP-1 to SECP-44), useful in the diagnosis, treatment and prevention of cardiovascular (e.g. atherosclerosis), autoimmune/inflammatory (e.g. allergies) and cell proliferative disorders.

XX PS Claim 5; Page 180-182; 195pp; English.

XX CC The invention relates to forty four human secreted proteins (referred to as SECP-1 to SECP-44) and the nucleic acids encoding them. Also included are a host cell transformed with the nucleic acid, a transgenic animal comprising the nucleic acid, an anti-SECP antibody, use of the SECP proteins in isolating agonists and antagonists of SECP activity and a method of isolating compounds which alter the expression of the SECP nucleic acid. The SECP polynucleotides and polypeptides are useful in the diagnosis, treatment and prevention of cardiovascular (e.g. atherosclerosis, hypertension, myocardial infarction), autoimmune/inflammatory (e.g. acquired immunodeficiency syndrome (AIDS), allergies, rheumatoid arthritis), cell proliferative (e.g. cancer), developmental (e.g. Duchenne and Becker muscular dystrophy), and neurological (e.g. epilepsy, Alzheimer's disease) disorders. Numerous other examples of each disorder are given in the specification. The present sequence is a cDNA encoding a SECP protein

XX SQ Sequence 4480 BP; 1092 A; 1152 C; 1102 G; 1134 T; 0 U; 0 Other;

Query Match 14.4%; Score 35.4; DB 6; Length 4480;
Best Local Similarity 63.5%; Pred. No. 2.7;
Matches 54; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 2 TCATTAACTTTCAGAGTACCTTTTATTTTCTTTAAGATTCCTGTTTATACACA 61
DB 3771 TCATTTTCCATCCAGGATAGATATATATTTCTTTGATTTTAAATATATATATA 3712

QY 62 GATTTTAAAGTTTACTCCTACTGCTG 86
DB 3711 TATATTATATGTACACACACCTG 3687

RESULT 11
ADE53917
ID ADE53917 standard; cDNA; 2264 BP.
XX AC ADE53917;
XX DT 29-JAN-2004 (first entry)
XX DE Human prostate cancer cDNA #264.
XX KW Human; prostate cancer; ss; cDNA combination; differential expression; gene.
XX OS Homo sapiens.
XX PN US2003190640-A1.
XX PD 09-OCT-2003.
XX PR 29-MAY-2002; 2002US-00252157.
XX PR 31-MAY-2001; 2001US-0295048P.
XX PA (FARI/) FARIS M.

```
PA (PEAR/) PEARSON C I.
XX
XX
PI Paris M, Pearson CI;
XX
XX WPI; 2003-831619/77.
XX
XX New combination comprising cDNAs that are differentially expressed in
PT prostate cancer, useful for diagnosing, treating or monitoring the
PT progression of treatment of prostate cancer.
XX
XX Claim 5; SEQ ID NO 264; 42pp; English.
XX
XX The invention relates to a combination comprising a number of cDNAs
XX expressed in prostate cancer. The invention also relates to a method for
XX detecting differential expression of one or more cDNAs in a sample
XX containing nucleic acids by hybridising a substrate with the nucleic
XX acids, thus forming one or more hybridisation complexes, detecting
XX hybridisation complex formation and comparing the complexes formed with
XX standard complexes, where differences between the standard and the sample
XX complex formation indicate differential expression of cDNAs in the
XX sample. The differential expression is diagnostic of prostate cancer. The
XX invention also relates to proteins and antibodies related to the cDNAs.
XX The combination is useful for diagnosing, treating or monitoring the
XX progression of treatment of prostate cancer. The antibodies are useful
XX for detecting prostate cancer. This sequence represents a human prostate
XX cancer cDNA of the invention.
XX
XX Sequence 2264 BP; 472 A; 629 C; 595 G; 517 T; 0 U; 61 Other;
SQ
Query Match 14.1%; Score 34.8; DB 9; Length 2264;
Best Local Similarity 54.3%; Pred. No. 3.3;
Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;
QY 25 TTTTATTTTCTTTAAGATTCCTGTGTTTATACACAGATTTTAAAGTTTACTCTCTACTGCG 84
DB 1452 TTCTGTTTCGTTTTTTTTTTTNNCTTTTCTTATATATATTTTGTGAATTCATTATTT 1511
QY 85 TGACCAAGTGAATTCCTCTCCAGTCACAGTCGCAACCTTACCCCACTCAACG 144
DB 1512 ATTTTAAATCTCTCTCTCTCCAGACACATGGCAGCTTATCTCGAATGCTGTG 1571
QY 145 AGAGTTT 151
DB 1572 ATCGTCT 1578
XX
XX RESULT 12
XX ABL03611/c
XX ID ABL03611 standard; cDNA; 2239 BP.
XX AC ABL03611;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 5315.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB59507.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 5315; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 2239 BP; 686 A; 496 C; 536 G; 521 T; 0 U; 0 Other;
SQ
Query Match 14.0%; Score 34.4; DB 4; Length 2239;
Best Local Similarity 63.1%; Pred. No. 4.3;
Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
QY 32 TTTCTTTAAGATTCCTGTGTTTATACACAGATTTTAAAGTTTACTCTCTACTGCGACCA 91
DB 1578 TTTCCTCACATCCATGTTGTCTATATACAGTGTCCATTGACTTTGTATCCTGACCTG 1519
QY 92 AGTGAATTCCTCTCCAGTCACA 115
DB 1518 TTGAAGATAATCCTCACTCCCA 1495
XX
XX RESULT 13
XX ABL03610/c
XX ID ABL03610 standard; cDNA; 4405 BP.
XX AC ABL03610;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster expressed polynucleotide SEQ ID NO 5312.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-656860/75.
XX
XX P-PSDB; ABB59507.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 5312; 21pp + Sequence Listing; English.
XX
XX
```


CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 4405 BP; 1275 A; 956 C; 1086 G; 1088 T; 0 U; 0 Other;
 Query Match 14.0%; Score 34.4; DB 4; Length 4405;
 Best Local Similarity 63.1%; Pred. No. 5.5; Mismatches 0; Indels 0; Gaps 0;
 Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
 QY 32 TTCTTTTAAGATTCCTGTTGTTTATACACAGATTTTAAGTTTACTCTCTGACCCCA 91
 DB 2744 TTTCCTCAGATCCCATGTTGTTCTATATACAGTGTTCCATTGACTTTGTGCTGACCTG 2685
 QY 92 AGTGAATTCCTTCCAGTCACA 115
 DB 2684 TTGAAGATAATCCTCAACTCCCA 2661
 RESULT 14
 ABL03530/c
 ID ABL03530 standard; cDNA; 12860 BP.
 AC ABL03530;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 5072.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ss.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 PD 27-SEP-2001.
 PF 23-MAR-2001; 2001WO-US009231.
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 DR P-PSDB; ABB59427.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions.
 XX Claim 1; SEQ ID NO 5072; 21pp + Sequence Listing; English.
 PS The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX
 SQ Sequence 12860 BP; 3591 A; 2899 C; 2995 G; 3375 T; 0 U; 0 Other;
 Query Match 14.0%; Score 34.4; DB 4; Length 12860;
 Best Local Similarity 63.1%; Pred. No. 8.1; Mismatches 0; Indels 0; Gaps 0;
 Matches 53; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
 QY 32 TTCTTTTAAGATTCCTGTTGTTTATACACAGATTTTAAGTTTACTCTCTGACCCCA 91
 DB 10389 TTTCCTCAGATCCCATGTTGTTCTATATACAGTGTTCCATTGACTTTGTGCTGACCTG 10330
 QY 92 AGTGAATTCCTTCCAGTCACA 115
 DB 10329 TTGAAGATAATCCTCAACTCCCA 10306
 RESULT 15
 AAK55128
 ID AAK55128 standard; cDNA; 388 BP.
 XX AAK55128;
 AC AAK55128;
 DT 06-NOV-2001 (first entry)
 DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:188.
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ss.
 OS Homo sapiens.
 PN WO200157182-A2.
 PD 09-AUG-2001.
 PF 17-JAN-2001; 2001WO-US001354.
 XX 31-JAN-2000; 2000US-0179065P.
 PR 04-FEB-2000; 2000US-0180628P.
 PR 24-FEB-2000; 2000US-0184664P.
 PR 02-MAR-2000; 2000US-0186350P.
 PR 16-MAR-2000; 2000US-0189874P.
 PR 17-MAR-2000; 2000US-0190076P.
 PR 18-APR-2000; 2000US-0198123P.
 PR 19-MAY-2000; 2000US-0205515P.
 PR 07-JUN-2000; 2000US-0209467P.
 PR 28-JUN-2000; 2000US-0214886P.
 PR 30-JUN-2000; 2000US-0215135P.
 PR 07-JUL-2000; 2000US-0216647P.
 PR 07-JUL-2000; 2000US-0216880P.
 PR 11-JUL-2000; 2000US-0217487P.
 PR 14-JUL-2000; 2000US-0217496P.
 PR 26-JUL-2000; 2000US-0220963P.
 PR 14-AUG-2000; 2000US-0224518P.
 PR 14-AUG-2000; 2000US-0224519P.
 PR 14-AUG-2000; 2000US-0225213P.
 PR 14-AUG-2000; 2000US-0225214P.
 PR 14-AUG-2000; 2000US-0225266P.
 PR 14-AUG-2000; 2000US-0225267P.
 PR 14-AUG-2000; 2000US-0225268P.
 PR 14-AUG-2000; 2000US-0225270P.
 PR 14-AUG-2000; 2000US-0225447P.
 PR 14-AUG-2000; 2000US-0225757P.
 PR 14-AUG-2000; 2000US-0225758P.
 PR 18-AUG-2000; 2000US-0226279P.
 PR 22-AUG-2000; 2000US-0226681P.
 PR 22-AUG-2000; 2000US-0226868P.
 PR 23-AUG-2000; 2000US-0227182P.
 PR 30-AUG-2000; 2000US-0227009P.
 PR 30-AUG-2000; 2000US-0228924P.

PR 01-SEP-2000; 2000US-0229287P.
 PR 01-SEP-2000; 2000US-0229343P.
 PR 01-SEP-2000; 2000US-0229344P.
 PR 01-SEP-2000; 2000US-0229345P.
 PR 05-SEP-2000; 2000US-0229509P.
 PR 05-SEP-2000; 2000US-0229513P.
 PR 06-SEP-2000; 2000US-0230437P.
 PR 06-SEP-2000; 2000US-0230438P.
 PR 08-SEP-2000; 2000US-0231242P.
 PR 08-SEP-2000; 2000US-0231243P.
 PR 08-SEP-2000; 2000US-0231244P.
 PR 08-SEP-2000; 2000US-0231413P.
 PR 08-SEP-2000; 2000US-0231414P.
 PR 08-SEP-2000; 2000US-0232080P.
 PR 08-SEP-2000; 2000US-0232081P.
 PR 12-SEP-2000; 2000US-0231968P.
 PR 14-SEP-2000; 2000US-0232397P.
 PR 14-SEP-2000; 2000US-0232398P.
 PR 14-SEP-2000; 2000US-0232399P.
 PR 14-SEP-2000; 2000US-0232400P.
 PR 14-SEP-2000; 2000US-0232401P.
 PR 14-SEP-2000; 2000US-0233063P.
 PR 14-SEP-2000; 2000US-0233064P.
 PR 14-SEP-2000; 2000US-0233065P.
 PR 21-SEP-2000; 2000US-0234223P.
 PR 21-SEP-2000; 2000US-0234274P.
 PR 25-SEP-2000; 2000US-0234997P.
 PR 25-SEP-2000; 2000US-0234998P.
 PR 26-SEP-2000; 2000US-0235484P.
 PR 27-SEP-2000; 2000US-0235834P.
 PR 27-SEP-2000; 2000US-0235836P.
 PR 29-SEP-2000; 2000US-0236327P.
 PR 29-SEP-2000; 2000US-0236367P.
 PR 29-SEP-2000; 2000US-0236368P.
 PR 29-SEP-2000; 2000US-0236369P.
 PR 29-SEP-2000; 2000US-0236370P.
 PR 02-OCT-2000; 2000US-0236802P.
 PR 02-OCT-2000; 2000US-0237037P.
 PR 02-OCT-2000; 2000US-0237038P.
 PR 02-OCT-2000; 2000US-0237039P.
 PR 02-OCT-2000; 2000US-0237040P.
 PR 13-OCT-2000; 2000US-0239935P.
 PR 13-OCT-2000; 2000US-0239937P.
 PR 20-OCT-2000; 2000US-0240960P.
 PR 20-OCT-2000; 2000US-0241221P.
 PR 20-OCT-2000; 2000US-0241785P.
 PR 20-OCT-2000; 2000US-0241786P.
 PR 20-OCT-2000; 2000US-0241787P.
 PR 20-OCT-2000; 2000US-0241808P.
 PR 20-OCT-2000; 2000US-0241809P.
 PR 20-OCT-2000; 2000US-0241826P.
 PR 01-NOV-2000; 2000US-024617P.
 PR 08-NOV-2000; 2000US-0246474P.
 PR 08-NOV-2000; 2000US-0246475P.
 PR 08-NOV-2000; 2000US-0246476P.
 PR 08-NOV-2000; 2000US-0246477P.
 PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.

PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249265P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251988P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251989P.
 PR 08-DEC-2000; 2000US-0251990P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-483426/52.
 XX P-PSDB; AAM82347.
 DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 DR useful for preventing, diagnosing and/or treating cancers and metastasis.
 XX Claim 1; SEQ ID NO 188; 3071pp + Sequence Listing; English.
 XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/haematopoietic-related diseases, especially
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/haematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
 CC represent sequences used in the exemplification of the present invention
 XX Sequence 388 BP; 76 A; 131 C; 73 G; 105 T; 0 U; 3 Other;
 SQ
 Query Match 13.9%; Score 34.2; DB 4; Length 388;
 Best Local Similarity 51.7%; Pred. No. 2.6; Indels 0; Gaps 0;
 Matches 78; Conservative 0; Mismatches 73;
 QY 1 GTCATTAACTTTCGAAGGATACCTTTTATTTCTTTCTTAAGATTCTCTGTTTATAC 60
 Db 205 GTCACAGTTCTCTCCACCTGTTCTGTTGTTTTTTTTTTCTTTCTTATAT 264
 QY 61 AGATTTTAAGTTTACTCTCTACTGCTGACCCAGTGAATTCCTTCTCCAGTCACAGTGC 120
 Db 265 ATATTTTGTGTAATTCATTATTTTATTTTAAATTCCTCTCTCTCTCCACACAAATGGC 324
 QY 121 AACCTCTACCCCACTGACAGAGTTT 151

Db 325 ACTGCTTATCTCGAAATGGTGTGATCGTCT 355

Search completed: June 10, 2004, 17:11:34
Job time : 364 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 10, 2004, 15:30:57 ; Search time 54 Seconds
(without alignments)
2528.110 Million cell updates/sec

Title: US-09-674-593-10
Perfect score: 246
Sequence: 1 gtcattacaccttgcaagga.....ttcatatcaacccacacta 246

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:
1: /cgn2_6/prodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/prodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	35.6	14.5	507	3	US-09-328-111-826
2	34	13.8	786431	4	Sequence 826, App
3	31.8	12.9	1885	4	Sequence 3, Appli
4	31.6	12.8	2337	4	Sequence 1162, Ap
5	31.6	12.8	13158	2	Sequence 3781, Ap
6	31.4	12.8	1017	4	Sequence 105, App
7	31.2	12.7	530	4	Sequence 42, Appl
8	31.2	12.7	530	4	Sequence 434, App
9	31.2	12.7	530	4	Sequence 434, App
10	31.2	12.7	530	4	Sequence 434, App
11	31.2	12.7	530	4	Sequence 434, App
12	31	12.6	704	3	Sequence 125, App
13	31	12.6	704	3	Sequence 125, App
14	31	12.6	1413	3	Sequence 1, Appli
15	31	12.6	3931	4	Sequence 342, App
16	31	12.6	5184	4	Sequence 484, App
17	30.6	12.4	550	4	Sequence 3345, App
18	30.6	12.4	741	4	Sequence 1894, Ap
19	30.2	12.3	9278	1	Sequence 9, Appli
20	30.2	12.3	9278	1	Sequence 9, Appli
21	30.2	12.3	9278	1	Sequence 9, Appli
22	30	12.2	129908	4	Sequence 1, Appli
23	29.8	12.1	8920	2	Sequence 1, Appli
24	29.8	12.1	8920	3	Sequence 1, Appli
25	29.6	12.0	1704	4	Sequence 263, App
26	29.6	12.0	3018	4	Sequence 40, Appl
27	29.6	12.0	3036	4	Sequence 38, Appl

C 28	29.6	12.0	3966	3	US-09-215-131-1	Sequence 1, Appli
C 29	29.6	12.0	3966	3	US-09-222-734-1	Sequence 1, Appli
C 30	29.4	12.0	324	4	US-09-107-532A-3015	Sequence 3015, Ap
C 31	29.4	12.0	704	3	US-08-235-836C-123	Sequence 123, App
C 32	29.4	12.0	704	3	US-08-235-836C-124	Sequence 124, App
C 33	29.2	11.9	337	4	US-08-956-171E-1348	Sequence 1348, Ap
C 34	28.8	11.7	495	4	US-09-543-691A-2486	Sequence 2486, Ap
C 35	28.6	11.6	503	4	US-09-621-976-15042	Sequence 15042, A
C 36	28.6	11.6	858	4	US-09-328-352-3863	Sequence 3863, Ap
C 37	28.6	11.6	1425	4	US-09-134-001C-404	Sequence 404, App
C 38	28.6	11.6	1882	2	US-08-867-970-1	Sequence 1, Appli
C 39	28.6	11.6	1882	3	US-09-326-217-1	Sequence 1, Appli
C 40	28.6	11.6	1882	4	US-09-732-020-1	Sequence 1, Appli
C 41	28.6	11.6	2946	3	US-08-968-563-6	Sequence 6, Appli
C 42	28.6	11.6	2946	3	US-08-969-683A-6	Sequence 6, Appli
C 43	28.6	11.6	2946	4	US-09-297-928-2	Sequence 2, Appli
C 44	28.4	11.5	56516	2	US-08-996-306-1	Sequence 1, Appli
C 45	28.4	11.5	56516	3	US-09-338-907-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-328-111-826
; Sequence 826, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Derti, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 826
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(507)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-826

Query Match 14.5%; Score 35.6; DB 3; Length 507;
Best Local Similarity 62.2%; Pred. No. 0.044;
Matches 56; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy	62	GATTTTAAGTTTACTCTCTACTGTGACCCCAAGTGAATTCCTTCTCCAGTCACAGTCA	121
Db	203	GATTTTAAATATATACCTCAGGACCAAGAAAAGTTAAGCAGCAGGTTCCA	262
Qy	122	ACCTTACCCCAACTGCAACGAGAGTTT	151
Db	263	AGTGTCTCTCTCCCACTTCAACAAGATGT	292

RESULT 2

```
US-09-751-389-3/c
; Sequence 3, Application US/09751389
; Patent No. 6630334
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001067
; CURRENT APPLICATION NUMBER: US/09/751,389
; CURRENT FILING DATE: 2001-01-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 786431
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(786431)
; OTHER INFORMATION: n = A,T,C or G
US-09-751-389-3

Query Match      13.8%; Score 34; DB 4; Length 786431;
Best Local Similarity 74.1%; Pred. No. 4.5;
Matches 43; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 16 AGGATACCTTTTATTTCTTTTAAAGATTCCTGTTGTTTATACACAGATTTTAAAGTTT 73
Db 691859 ATGGATACATTTTATAATATTTAAATTTTATCTTTTAAACACACACATTTAAGATT 691802

RESULT 3
US-09-023-655-1162
; Sequence 1162, Application US/09023655
; Patent No. 6607879
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
; TITLE OF INVENTION: EXPRESSION
; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1162:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1885 base pairs
```

```
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: g187386
US-09-023-655-1162

Query Match      12.9%; Score 31.8; DB 4; Length 1885;
Best Local Similarity 53.7%; Pred. No. 1.4;
Matches 66; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 29 TATTTCCTTTAAGATTCCTGTTGTTTATACACAGATTTTAAAGTTTACTCCTACTGCTGAC 88
Db 1426 TGTCTTTTATTTTACTTTTATTTTAAAGCAAAATTTGTTGTTTCTCCCTCCCTCCCTCC 1485

QY 89 CCAAGTGAATTCCTTCCTCCAGTCACAGTCACACCTCTACCCCACTGCAACGAGAG 148
Db 1486 TCCCACAGATCCCATCTCAAAATCTCTGTATACCACTCCCAAGTCGAGGAGAG 1545

QY 149 TTT 151
Db 1546 CTT 1548

RESULT 4
US-09-328-352-3781
; Sequence 3781, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3781
; LENGTH: 2337
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3781

Query Match      12.8%; Score 31.6; DB 4; Length 2337;
Best Local Similarity 51.4%; Pred. No. 1.8;
Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 13 TGCACGATACCTTTTATTTTCTTTTAAAGTTCCTGTTGTTTATACACAGATTTTAAAGTT 72
Db 1281 TGAGACGGAAGGTTTGTGATCAGCTTCAAGACTTTGAAGGTTTGAAGGTTTGAAGGTTT 1340

QY 73 TACTCCTACTGCTGACCCCAAGTGAATTCCTTCTCCAGTCACAGTGTCAACCTTACCC 132
Db 1341 TAGCCGTGATGGTGATATTAAAGAACTACACCGTGGCTCAACAGTTTATAGCTTTGCTTA 1400

QY 133 CCAACTGCAACGAGAGTTTGA 154
Db 1401 CCATGTGCATACAGAGGTTGA 1422

RESULT 5
US-08-687-080-105
; Sequence 105, Application US/08687080
; Patent No. 5965427
; GENERAL INFORMATION:
; APPLICANT: Gregory Dolganov
; TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof
; NUMBER OF SEQUENCES: 175
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
```

COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,080
FILING DATE: 17-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/592,126
FILING DATE: 26-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0111.30
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 105:
SEQUENCE CHARACTERISTICS:
LENGTH: 13158 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
INDIVIDUAL ISOLATE: 5' END OF INTRON 21 OF RAD50 GENOMIC
SEQUENCE
INDIVIDUAL ISOLATE: SEQUENCE
US-08-687-080-105

Query Match 12.8%; Score 31.6; DB 2; Length 13158;
Best Local Similarity 69.4%; Pred. No. 4;
Matches 43; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 25 TTTTATTTCTTTAAGATTCCTGTTTATACACAGATTTTAAAGTTTACTCCTACTGC 84
DB 4363 TTTATATTCCTTTTACAGATTTGTTTATTTTAAATTCACATTTTGAGGAGTCGTGCTTCTGC 4422
QY 85 TG 86
DB 4423 TG 4424

RESULT 6
US-09-601-198-42
Sequence 42, Application US/09601198
Patent No. 6531583
GENERAL INFORMATION:
APPLICANT: Cassell, Gail H.
APPLICANT: Chen, Ellison Y.
APPLICANT: Glass, Jennifer S.
APPLICANT: Glass, John I.
APPLICANT: Reiner, Cheryl R.
APPLICANT: Leikowitz, Elliot
TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREA PLASMA
FILE REFERENCE: UAB-13452/22
CURRENT APPLICATION NUMBER: US/09/601,198
CURRENT FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/073,189
PRIOR FILING DATE: 1998-01-30
NUMBER OF SEQ ID NOS: 181
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 42
LENGTH: 1017
TYPE: DNA
ORGANISM: Ureaplasma urealyticum
US-09-601-198-42

Query Match 12.8%; Score 31.4; DB 4; Length 1017;
Best Local Similarity 53.7%; Pred. No. 1.4;
Matches 65; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
QY 7 AACCTTGCAGGATACCTTTTATTTCTTTAAGATTCCTGTTTATACACAGATTT 66
DB 589 AATATCTCTGTATAAATCTTCATTGCTTTGATTTTCGTTTATAGGCTCAGGTC 648
QY 67 TAAGTTTACTCCTACTGCTGACCAAGTGAATTCCTCTCCAGTCACAGTGTCAACCTC 126
DB 649 AACTTCTTCATGTAATGTTCAATAGGTGAGTCAATGTTGTTAAACAATATCATCTTC 708
QY 127 T 127
DB 709 T 709

RESULT 7
US-09-389-681-434/c
Sequence 434, Application US/09389681A
Patent No. 6518237
GENERAL INFORMATION:
APPLICANT: Yuqui, Jiang
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Xu, Jiangchun
TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
DIAGNOSIS OF BREAST CANCER AND METHODS FOR THEIR USE
FILE REFERENCE: 210121.470C3
CURRENT APPLICATION NUMBER: US/09/389,681A
CURRENT FILING DATE: 1999-09-02
NUMBER OF SEQ ID NOS: 463
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 434
LENGTH: 530
TYPE: DNA
ORGANISM: Homo sapiens
US-09-389-681-434

Query Match 12.7%; Score 31.2; DB 4; Length 530;
Best Local Similarity 60.7%; Pred. No. 1.2;
Matches 51; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 24 CTTTATTTCTTTAAGATTCCTGTTTATACACAGATTTTAAAGTTTACTCCTACTG 83
DB 466 CTTTATTTCTTTAAGATTTTAACTTTACTTTTCTCATAGAGAGTCTTCTGTCATTATCTCACT 407
QY 84 CTGACCCAGTGAATTCCTTCTC 107
DB 406 CTCACAAAATCAGAAATCATCTC 383

RESULT 8
US-09-620-405B-434/c
Sequence 434, Application US/09620405B
Patent No. 6528054
GENERAL INFORMATION:
APPLICANT: Jiang, Yuqui
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Xu, Jiangchun
APPLICANT: Harlocker, Susan L.
APPLICANT: Hepler, William T.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
DIAGNOSIS OF BREAST CANCER
FILE REFERENCE: 210121.470C8
CURRENT APPLICATION NUMBER: US/09/620,405B
CURRENT FILING DATE: 2000-07-20
NUMBER OF SEQ ID NOS: 495
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 434
LENGTH: 530

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-620-405B-434

Query Match      12.7%; Score 31.2; DB 4; Length 530;
Best Local Similarity 60.7%; Pred. No. 1.2;
Matches 51; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 24 CTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGATTTTAAAGTTTACTCCTACTG 83
Db 466 CTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGATTTTAAAGTTTACTCCTACT 407

QY 84 CTGACCCCAAGTGAATTCCTTCTC 107
Db 406 CTCACAAAATCAGATCATCCTC 383

RESULT 9
US-09-433-826B-434/c
; Sequence 434, Application US/09433826B
; Patent No. 6579973
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqui
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS FOR THE TREATMENT AND
; FILE REFERENCE: 210121.470C4
; CURRENT APPLICATION NUMBER: US/09/433,826B
; CURRENT FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 474
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 434
; LENGTH: 530
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-433-826B-434

Query Match      12.7%; Score 31.2; DB 4; Length 530;
Best Local Similarity 60.7%; Pred. No. 1.2;
Matches 51; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 24 CTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGATTTTAAAGTTTACTCCTACTG 83
Db 466 CTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGATTTTAAAGTTTACTCCTACT 407

QY 84 CTGACCCCAAGTGAATTCCTTCTC 107
Db 406 CTCACAAAATCAGATCATCCTC 383

RESULT 10
US-09-604-287A-434/c
; Sequence 434, Application US/09604287A
; Patent No. 6586572
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqui
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C7
; CURRENT APPLICATION NUMBER: US/09/604,287A
; CURRENT FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 489
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 434
; LENGTH: 530
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-604-287A-434

Query Match      12.7%; Score 31.2; DB 4; Length 530;
Best Local Similarity 60.7%; Pred. No. 1.2;
Matches 51; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 24 CTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGATTTTAAAGTTTACTCCTACTG 83
Db 466 CTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGATTTTAAAGTTTACTCCTACT 407

QY 84 CTGACCCCAAGTGAATTCCTTCTC 107
Db 406 CTCACAAAATCAGATCATCCTC 383

RESULT 11
US-09-834-759-434/c
; Sequence 434, Application US/09834759
; Patent No. 6680197
; GENERAL INFORMATION:
; APPLICANT: Jiang, Yuqui
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Hepler, William T.
; APPLICANT: Henderson, Robert A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.470C9
; CURRENT APPLICATION NUMBER: US/09/834,759
; CURRENT FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 547
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 434
; LENGTH: 530
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-834-759-434

Query Match      12.7%; Score 31.2; DB 4; Length 530;
Best Local Similarity 60.7%; Pred. No. 1.2;
Matches 51; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 24 CTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGATTTTAAAGTTTACTCCTACTG 83
Db 466 CTTTATTTTCTTTAAGATTCCTGTTGTTTATACACAGATTTTAAAGTTTACTCCTACT 407

QY 84 CTGACCCCAAGTGAATTCCTTCTC 107
Db 406 CTCACAAAATCAGATCATCCTC 383

RESULT 12
US-08-235-836C-125/c
; Sequence 125, Application US/08235836C
; Patent No. 6248562
; GENERAL INFORMATION:
; APPLICANT: Dunn, John J.
; APPLICANT: Luft, Benjamin J.
; TITLE OF INVENTION: No. 6248562el Chimeric Proteins Comprising
; TITLE OF INVENTION: Borrelia Polypeptides and Uses Therefor
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESS: Brookhaven National Laboratory
; STREET:
; CITY: Upton
; STATE: NY
; COUNTRY: USA
; ZIP: 11973
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/235,836C
FILING DATE: 29-APR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/148,191
FILING DATE: 01-11-93
ATTORNEY/AGENT INFORMATION:
NAME: Bogosian, Margaret C.
REGISTRATION NUMBER: 25,324
REFERENCE/DOCKET NUMBER: BNL93-28A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 282-7338
TELEFAX: (516) 282-3729
INFORMATION FOR SEQ ID NO: 125:
SEQUENCE CHARACTERISTICS:
LENGTH: 704 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-235-836C-125

Query Match 12.6%; Score 31; DB 3; Length 704;
Best Local Similarity 52.8%; Pred. No. 1.6;
Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 2 TCATTAACTTTGCAAGGATACCTTTTATTTCTTTTAAAGATTCCTGTGTTTATACACA 61
Db 463 TCGTTCTGCTTCATCAGATCAGATAGTGTGTTAGTTGCTTCTTCTTGTGTTACGT 404

QY 62 GATTTTAAGTTTACTCTACTGCTGACCCCAAGTGAATTCCTTCCAGTCACAGTGTCA 121
Db 403 TTTTGCAATATTTTGTCTACTGCTTTAGTAGAGTTATATGCTTCTTTCGCAACATTAGCC 344

QY 122 ACCTCTA 128
Db 343 GCAACTA 337

RESULT 13
US-08-235-836C-126/c
; Sequence 126, Application US/08235836C
; Patent No. 6248562
; GENERAL INFORMATION:
; APPLICANT: Dunn, John J.
; APPLICANT: Luft, Benjamin J.
; TITLE OF INVENTION: No. 6248562a1 Chimeric Proteins Comprising
; TITLE OF INVENTION: Borrelia Polypeptides and Uses Therefor
; NUMBER OF SEQUENCES: 144
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brookhaven National Laboratory
; STREET:
; CITY: Upton
; STATE: NY
; COUNTRY: USA
; ZIP: 11973
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/235,836C
; FILING DATE: 29-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,191
; FILING DATE: 01-11-93
; ATTORNEY/AGENT INFORMATION:

NAME: Bogosian, Margaret C.
REGISTRATION NUMBER: 25,324
REFERENCE/DOCKET NUMBER: BNL93-28A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 282-7338
TELEFAX: (516) 282-3729
INFORMATION FOR SEQ ID NO: 126:
SEQUENCE CHARACTERISTICS:
LENGTH: 704 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-235-836C-126

Query Match 12.6%; Score 31; DB 3; Length 704;
Best Local Similarity 52.8%; Pred. No. 1.6;
Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 2 TCATTAACTTTGCAAGGATACCTTTTATTTCTTTTAAAGATTCCTGTGTTTATACACA 61
Db 463 TCGTTCTGCTTCATCAGATCAGATAGTGTGTTAGTTGCTTCTTCTTGTGTTACGT 404

QY 62 GATTTTAAGTTTACTCTACTGCTGACCCCAAGTGAATTCCTTCCAGTCACAGTGTCA 121
Db 403 TTTTGCAATATTTTGTCTACTGCTTTAGTAGAGTTATATGCTTCTTTCGCAACATTAGCC 344

QY 122 ACCTCTA 128
Db 343 GCAACTA 337

RESULT 14
US-09-342-459-1/c
; Sequence 1, Application US/09342459
; Patent No. 6251633
; GENERAL INFORMATION:
; APPLICANT: Yan, Kang
; APPLICANT: Pearce, Kenneth
; TITLE OF INVENTION: FtsA
; FILE REFERENCE: GM10229
; CURRENT APPLICATION NUMBER: US/09/342,459
; CURRENT FILING DATE: 1999-06-29
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1413
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; US-09-342-459-1

Query Match 12.6%; Score 31; DB 3; Length 1413;
Best Local Similarity 56.3%; Pred. No. 2.2;
Matches 58; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 2 TCATTAACTTTGCAAGGATACCTTTTATTTCTTTTAAAGATTCCTGTGTTTATACACA 61
Db 1364 TCTTTATCTTTGTTCTGACATGTTCAATGTCGTTTATGTTCTCTGATGATTAICTTCT 1305

QY 62 GATTTTAAGTTTACTCTACTGCTGACCCCAAGTGAATTCCTT 104
Db 1304 GATTGATAAATTTCTTCTCATCTGTTGACTTACTTCAATTTCAT 1262

RESULT 15
US-08-956-171E-342/c
; Sequence 342, Application US/08956171E
; Patent No. 6593114
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; Gil H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash

Mon Jun 14 07:43:39 2004

Michael R. Fannon
 TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
 NUMBER OF SEQUENCES: 5256
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/956,171E
 FILING DATE: 20-Oct-1997
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 60/009,861
 FILING DATE: January 5, 1996
 APPLICATION NUMBER: 08/781,986
 FILING DATE: January 3, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Mark J. Hyman
 REGISTRATION NUMBER: 46,789
 REFERENCE/DOCKET NUMBER: PB248P1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (240) 314-1224
 TELEFAX: (301) 309-8439
 INFORMATION FOR SEQ ID NO: 342:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3931 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 342:
 US-08-956-171E-342

Query Match 12.6%; Score 31; DB 4; Length 3931;
 Best Local Similarity 56.3%; Pred. No. 3.6;
 Matches 58; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
 QY 2 TCATTAACTTTCGAAGGATACCTTTTATTCTTTTAAAGATTCCTGTTGTTTATACACA 61
 Db 500 TCATTATCTTTGCTTGAACATGTCATGATTCGTTTATGTTCTCTGATGATTCTTCT 441
 QY 62 GATTTTAAGTTTACTCTACTGCTGACCCCAAGTGAATTCCTT 104
 Db 440 GATTGATAAATTTCTTCATCTGTTGACTCTACTCTACTTCATTTTCAT 398

Search completed: June 10, 2004, 17:01:40
 Job time : 59 secs